



**KORONER ARTER HASTALIĞINDA KLİNİK
SONLANIMLARIN ÖNGÖRÜLMESİNDE LOJİSTİK
REGRESYON ANALİZİ, CART VE RASTGELE ORMAN
ALGORİTMALARININ KARŞILAŞTIRILMASI**

Selim TOPCU
Biyoistatistik Anabilim Dalı

Tez Danışmanı
Prof. Dr. Mehmet Suphi ÖZÇOMAK

Doktora Tezi-2023

**T.C.
ATATÜRK ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**KORONER ARTER HASTALIĞINDA KLİNİK
SONLANIMLARIN ÖNGÖRÜLMESİNDE LOJİSTİK
REGRESYON ANALİZİ, CART VE RASTGELE ORMAN
ALGORİTMALARININ KARŞILAŞTIRILMASI**

Selim TOPCU

**Biyoistatistik Anabilim Dalı
Doktora Tezi**

**Tez Danışmanı
Prof. Dr. Mehmet Suphi ÖZÇOMAK**

**ERZURUM
2023**

İÇİNDEKİLER

İÇİNDEKİLER	I
TEŞEKKÜR	IV
ÖZET	V
ABSTRACT.....	VI
ŞEKİLLER DİZİNİ	IX
TABLolar DİZİNİ.....	XI
1. GİRİŞ	1
2. GENEL BİLGİLER.....	3
2.1. Karar Ağaçları.....	3
2.1.1. Karar Ağacı Türleri.....	8
2.1.2. Karar Ağaçlarının Üstünlükleri	10
2.1.3. Sınırlılıkları.....	11
2.2. Lojistik Regresyon Analizi	12
2.3. Sınıflandırma ve Regresyon Ağacı (CART).....	15
2.3.1. Gini Ayırma Kriteri	17
2.3.2. Twoing Ayırma Kriteri	18
2.3.3. En Küçük Kareler Sapması (LSD).....	19
2.4. Rastgele Orman Algoritması	20
2.4.1. Rastgele Orman Algoritması Akış Şeması	22
2.5. Stres Ekokardiyografi	24
2.5.1. Stres Ekokardiyografi Patofizyoloji.....	25
2.5.2. Dobutamin Stres Ekokardiyografi	26
2.5.2.1. Dobutamin Stres Ekokardiyografi Protokolü	27
2.5.2.2. Dobutamin Stres Ekokardiyografi Yanıtları	28

2.5.2.3. Dobutamin İnfüzyonunu Durdurma Kriterleri.....	29
2.5.2.4. Hasta Monitörizasyonu	29
2.5.2.5. Dobutamin Stres Ekokardiyografi İçin Kontrendikasyonlar	30
3. MATERYAL VE METOT.....	31
3.1. Araştırma Türü.....	31
3.2. Araştırmanın Yapıldığı Yer ve Zaman	31
3.3. Araştırma Popülasyonu.....	31
3.4. Dobutamin Stres Ekokardiyografi	31
3.5. Araştırma Değişkenlerinin Tanımlanması	33
3.6. İstatistiksel Analizler	34
3.7. Araştırmada Kullanılan Analizler ve Modellerin Oluşturulması.....	34
3.8. Model Performanslarının Kıyaslanması	34
4. BULGULAR.....	35
4.1. Bağımsız Değişkenler	35
4.1.1. Kategorik Değişkenler	35
4.1.1.1. Cinsiyet	35
4.1.1.2. Hipertansiyon.....	36
4.1.1.3. Diyabetes Mellitus	37
4.1.1.4. Koroner Arter Hastalığı	38
4.1.1.5. İstirahat Duvar Hareket Kusuru	39
4.1.1.6. Stres Testi Pozitifliği	40
4.1.1.7. Göğüs Ağrısı Varlığı.....	41
4.1.1.8. Yeni MI Varlığı	42
4.1.1.9. Ölüm	43
4.1.1.10. Yeni PTCA İhtiyacı	44

4.1.1.11. Yeni CABG İhtiyacı	45
4.1.1.12. Klinik Sonlanım- Girişimsel Sonlanım.....	46
4.1.1.13. Kardiyak Olay.....	47
4.1.2. Sayısal Değişkenler.....	49
4.1.2.1. Sistolik Kan Basıncı.....	49
4.1.2.2. Dobutamin Dozu.....	50
4.1.2.3. Bazal Ejeksiyon Fraksiyonu	50
4.1.2.4. Dobutamin Ejeksiyon Fraksiyonu	51
4.2. Değişkenlerin Önemlilik Sırası.....	52
4.3. Model Performansların Karşılaştırılması.....	54
5. TARTIŞMA.....	59
6. SONUÇ VE ÖNERİLER.....	65
KAYNAKLAR	67
EKLER	79
EK-1. ÖZGEÇMİŞ	79
EK-2. ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU	80
EK-3. ETİK KURUL ONAY FORMU	81
EK-4. TEZ ADI DEĞİŞİKLİĞİ FORMU	83

TEŞEKKÜR

Eğitim süreci boyunca yardımlarını ve yol göstericiliğini her zaman gösteren danışman hocam, Prof. Dr. Mehmet Suphi ÖZÇOMAK şükranlarımı sunarım. Biyoistatistik ilgimin temel kaynağı olan ve bu süreçte beni her zaman motive eden ve destekleyen değerli hocam Prof. Dr. İbrahim Halil TANBOĞA' ya, eğitim süreci boyunca yardımlarını esirgemeyen değerli arkadaşım Doç. Dr. Uğur AKSU' ya teşekkürlerimi sunuyorum. Ayrıca bu uzun süreçte tez ara rapor sunumlarında kıymetli katkıları ile tezin son halini almasında yardımları için Prof. Dr. Ömer ALKAN ve Doç. Dr. Hakan EYGÜ hocalarıma şükranlarımı sunuyorum.

Son olarak, bu süreçte zamanlarını çaldığım ve buna rağmen beni hep destekleyen eşime ve çocuklarıma sonsuz teşekkürler...

SELİM TOPCU

ÖZET

Koroner Arter Hastalığında Klinik Sonlanımların Öngörülmesinde Lojistik Regresyon Analizi, CART ve Rastgele Orman Algoritmalarının Karşılaştırılması

Amaç: Biz bu çalışmamızda klinik tahmin modeli olarak kullanılabilecek çok değişkenli regresyon ağaçlarından logistik regresyon analizi, sınıflama ve regresyon ağacı (CART) ve rastgele orman(random forest) yöntemleri tartışıldı ve örnek bir veri seti ile bu ağaçların performanslarını karşılaştırılmıştır.

Materyal ve Metot: Çalışmaya Mart 1991 ile Mart 1996 arasında, UCLA Tıp Fakültesi Yetişkin Kardiyak Görüntüleme ve Hemodinami laboratuvarına sevk edilen 1138 hasta dahil edilmiştir. Ayrıca, 12 aylık takibi olmayan 376 hasta da çalışmaya dahil edilmedi. Nihai çalışma popülasyonu 558 kişiden oluşmaktadır. Tüm hastalarda, DSE sonrası takip eden 12 ay içinde aşağıdakilerden herhangi birinin meydana gelip gelmediğini belirlemek için tıbbi kayıt gözden geçirilmiştir: “herhangi bir nedene bağlı ölüm, yeniden MI, takipte PTCA veya CABG ihtiyacı”. Analize yalnızca en az 12 aylık takibi olan veya ilk 12 ay içinde “herhangi bir nedene bağlı ölüm, yeniden MI, takipte PTCA veya CABG ihtiyacı” olarak tanımlanan bir kardiyak olayı olan hastalar dahil edilmiştir. Rastgele orman, CART ve logistik regresyon modellerinin dobutamin stres ekokardiyografi sonuçlarını yorumlama güçleri ve model performansları karşılaştırılmıştır.

Bulgular: Çalışmamızda; 8 prediktör her üç regresyon modeline dahil edilerek modellerin güçleri ile ölüm, MI, bypass ve PTCA öngördürücülükleri karşılaştırılmıştır. Model performanslarını karşılaştırılması için kullanılan birçok ölçüt bulunmaktadır. Biz tezimizde performans karşılaştırması için R2, Brier skoru, ölçekli brier, AUC (C istatistiği), kalibrasyon intercept, kalibrasyon slope ve ECI (estimated kalibrasyon indeksi) kullandık. R2 değeri CART için 0.283’dir ve en yüksektir. Rastgele orman için R2 değeri 0.118’dir. Brier skoru en düşük olan CART analizidir ve en iyi performans göstergesi ona aittir. AUC değeri en yüksek olan CART analizidir. AUC CART analizi 0.785’dir ve logistik regresyon ve rastgele orman için sırasıyla 0.772; 0.704’tür.

Sonuç: Her üç model de stres ekokardiyografi sonuçlarına göre klinik tahmin yapmada iyi düzeyde performans göstermiştir. CART en iyi model performansı sağlarken rastgele orman en düşük düzeyde model performansı sağlamıştır.

Anahtar Kelimeler: Klinik tahmin modelleri, koroner arter hastalığı, regresyon ağaçları

ABSTRACT

Comparison Of Logistic Regression Analysis, CART And Random Forest Algorithms For Predicting Clinical Outcomes In Patients With Coronary Artery Disease

Aim: In this study, we will try to discuss logistic regression analysis, classification and regression tree (CART) and random forest (RF) methods from multivariate regression trees that can be used as clinical prediction models, and we will try to compare the performances of these trees with a sample data set.

Material and method: Between March 1991 and March 1996, 1138 patients referred to the UCLA School of Medicine Adult Cardiac Imaging and Hemodynamics laboratory were recruited. In addition, 376 patients who did not have a 12-month follow-up were excluded from the study. The final study population consisted of 558 individuals. In all patients, the medical record was reviewed to determine if any of the following occurred within 12 months after DSE: “death from any cause, re-MI, need for PTCA or CABG at follow-up”. Only patients with at least 12 months of follow-up or a cardiac event within the first 12 months defined as “death from any cause, re-MI, need for PTCA or CABG at follow-up” were included in the analysis. The power to interpret dobutamine stress echocardiography results and model performances of the RF, CART and logistic regression models were compared.

Results: In our study; Eight predictors were included in all three regression models, and the power of the models and their predictors of death, MI, bypass and PTCA were compared. There are many criteria used to compare model performances. We used R², Brier score, scaled brier, AUC (C statistic), calibration intercept, calibration slope and ECI (estimated calibration index) for performance comparison in our thesis. The R² value is 0.283 for CART and is the highest. The R² value for the rastgele orman is 0.118. The CART analysis has the lowest Brier score and has the best performance indicator. It is the CART analysis with the highest AUC value. AUC CART analysis is 0.785 and 0.772 and 0.704 for logistic regression and random forest, respectively.

Conclusion: All three models performed well in making clinical predictions based on stress echocardiography results. CART provided the best model performance, while random forest provided the lowest model performance.

Key Words: Clinical prediction models, coronary artery disease, regression trees

SİMGELER VE KISALTMALAR DİZİNİ

ACEİ	:	Anjiyotensin konverting enzim inhibitörü
AF	:	Atrial fibrilasyon
ARB	:	Anjiyotensin reseptör blokeri
AUC	:	Area under curve(eğri altında kalan alan)
CABG	:	Koroner arter bypas grefti
CART	:	Classification and regression tree
CHAID	:	Chi-Squared Automatic Interaction Detector: Otomatik Ki-Kare Etkileşim Belirleme),
DCA	:	Desicion curve analysis
DM	:	Diabetes mellitus
DSE	:	Dobutamin stres ekokardiyografisi
EAE	:	Avrupa ekokardiyografi cemiyeti
ECI	:	Estimated kalibrasyon indeksi
EF	:	Ejeksiyon fraksiyonu
EKG	:	Elektrokardiyografi
EKO	:	Ekokardiyografi
HT	:	Hipertansiyon
IV	:	İntravenöze
KAH	:	Koroner arter hastalığı
LR	:	Logistik regresyon
LRA	:	Logistik regresyon analizi
LSD	:	En küçük kareler yöntemi
LVEF	:	Sol ventrikül ejeksiyon fraksiyonu

MARS	:	Multivariate Adaptive Regression Splines: Çok Değişkenli Uyumlu Regresyon Uzanımları),
MI	:	Miyokard infarktüsü
PCA	:	Temel bileşen analizi
PET	:	Pozitron emsiyon tomografisi
PTCA	:	Perkütan translüminal koroner anjiyoplast
QUEST	:	Quick, Unbiased, Efficient Statistical Tree: Hızlı, Yansız, Etkin İstatistiksel Ağaç
RF	:	Rastgele orman(Rastgele orman)
SKB	:	Sistolik kan basıncı
SLIQ	:	Supervised Learning in Quest),
SPRINT	:	Scalable Parallelizable Induction of Decision Trees)

ŞEKİLLER DİZİNİ

<u>Şekil No</u>	<u>Sayfa No</u>
Şekil 2.1. Basit bir karar ağacı örneği	5
Şekil 2.2. Karar ağacı yapısı.....	7
Şekil 2.3. Rastgele orman akış şeması.....	21
Şekil 2.4. Karar ağacı ve rastgele orman karşılaştırılması	24
Şekil 2.5. Standard dobutamin stres ekokardiyografi protokolü	27
Şekil 4.1. Cinsiyet değişkeni için pasta grafiği	36
Şekil 4.2. Hipertansiyon değişkeni için pasta grafiği	37
Şekil 4.3. Diyabetes mellitus değişkeni için pasta grafiği.....	38
Şekil 4.4. Koroner arter hastalığı varlığı değişkeni için pasta grafiği	39
Şekil 4.5. İstirahat duvar hareket kusuru değişkeni için pasta grafiği.....	40
Şekil 4.6. Stres testi değişkeni değişkeni için pasta grafiği.....	41
Şekil 4.7. Göğüs ağrısı varlığı değişkeni için pasta grafiği	42
Şekil 4.8. Yeni MI varlığı değişkeni için pasta grafiği.....	43
Şekil 4.9. Hipertansiyon değişkeni için pasta grafiği	44
Şekil 4.10. Ölüm değişkeni için pasta grafiği.....	45
Şekil 4.11. Yeni CABG değişkeni için pasta grafiği	46
Şekil 4.12. Klinik sonlanım ve girişimsel sonlanım değişkenleri sütun grafiği	47
Şekil 4.13. Kardiyak olay varlığı için sütun grafiği.....	48
Şekil 4.14. Sistolik kan basıncı için verilen dağılım grafiği.....	49
Şekil 4.15. Dobutamin dozu için verilen dağılım grafiği	50
Şekil 4.16. Bazal EF için verilen dağılım grafiği	51
Şekil 4.17. Dobutamin EF için verilen dağılım grafiği	51
Şekil 4.18. Logistik regresyon analizi için değişkenleri önemlilik sırası grafiği	52

Şekil 4.19. CART için değişkenleri önemlilik sırası grafiği	53
Şekil 4.20. Rastgele orman için değişkenleri önemlilik sırası grafiği	54
Şekil 4.21. Logistik regresyon analizi için kalibrasyon eğrisi.....	55
Şekil 4.22. Rastgele orman algoritması için kalibrasyon eğrisi	56
Şekil 4.23. CART için kalibrasyon eğrisi.....	57
Şekil 4.24. CART, logistik regresyon analizi ve Rastgele Orman için DCA eğrisi	57
Şekil 4.25. Klinik sonlanım için CART ağacı	58



TABLÖLAR DİZİNİ

<u>Tablo No</u>	<u>Sayfa No</u>
Tablo 2.1. Veri Madenciliğinde Kullanılan Modeller.....	4
Tablo 2.2. Karar ağacı düğümleri	7
Tablo 2.3. Dobutamin stres ekokardiyografinin yorumlanması.....	29
Tablo 3.1. Araştırma değişkenlerinin tanımlanması	33
Tablo 4.1. Hastaların Temel Demografik, Klinik ve Laboratuvar Bulguları.....	35
Tablo 4.2. Cinsiyet değişkeni için sayı ve yüzdeler	35
Tablo 4.3. Hipertansiyon değişkeni için sayı ve yüzdeler	36
Tablo 4.4. Diyabetes mellitus değişkeni için sayı ve yüzdeler	37
Tablo 4.5. Koroner arter hastalığı değişkeni için sayı ve yüzdeler	38
Tablo 4.6. İstirahat duvar hareket kusuru değişkeni için sayı ve yüzdeler	39
Tablo 4.7. Stres testi değişkeni için sayı ve yüzdeler	40
Tablo 4.8. Göğüs ağrısı varlığı değişkeni için sayı ve yüzdeler	41
Tablo 4.9. Yeni MI varlığı değişkeni için sayı ve yüzdeler	42
Tablo 4.10. Ölüm değişkeni için sayı ve yüzdeler	43
Tablo 4.11. Yeni PTCA ihtiyacı değişkeni için sayı ve yüzdeler	44
Tablo 4.12. Yeni CABG değişkeni için sayı ve yüzdeler	45
Tablo 4.13. Klinik sonlanım ve girişimsel sonlanım değişkenleri için sayı ve yüzdeler.....	46
Tablo 4.14. Kardiyak olay gelişmesi değişkeni için sayı ve yüzdeler	47
Tablo 4.15. Kardiyak olan gelişen ve gelişmeyen gruplarda değişkenlerin karşılaştırılması	48
Tablo 4.16. Sayısal değişkenlere ait veriler	49
Tablo 4.17. Değişkenlerin önemlilik sırası	52
Tablo 4.18. CART, LRA ve RF için model performans ölçütlerinin karşılaştırılması ..	55

1. GİRİŞ

Biyomedikal arařtırmalarda istatistik bilgisi zorunludur ve arařtırmanın ayrılmaz bir parçasıdır. İstatistik, ampirik verilerin toplanması, analiz edilmesi, yorumlanması ve rapor haline getirerek sunulması için yöntemler oluşturmak ve arařtırmakla ilgilenen bir bilim dalıdır. İstatistik disiplinler arası bir alan olup tüm bilimsel alanlarda uygulanabilirliđi vardır. Kanıta dayalı tıp çağında, biyomedikal gözlem ve deneyler tasarlamak ve yürütmek, bunlardan elde edilen verileri sunmak ve sonuçları yorumlamak istatistik uygulamaları olmadan mümkün değildir. İstatistik uygulamalarının sađlık alanında kullanımı ile biyoistatistik bilimi oluşmuştur. Biyoistatistik, bilimsel olarak veri toplamanın ve toplanan verileri özetlemenin yanı sıra, tüm gözlemsel ve deneysel çalışmalardan elde edilen arařtırma sorularının hipotezlerini test etmek için kullanılır. Bilim adamları veya arařtırmacılar, bir hastalığın hedef popölasyonu etkileme olasılıđını belirlemek için belirli bir veri kümesi için biyoistatistik ve olasılık teorisini birleřtirir. Bu nedenle, istatistiksel yöntemler mevcut verileri analiz etmede kullanıldıđı gibi, gelecekte oluşabilecek olumlu veya olumsuz durumları tahmin etmek için de kullanılabilir.¹

Bilindiđi gibi istatistik alanındaki iki temel fikir mevcuttur. Bunların birincisi belirsizliktir. Diđeri ise deđişkenliktir. Sonucun çođunlukla belirli olmadığı bilimde karřılařılan birçok durum vardır. Bu durumların bazısında belirsizlik, beklenen sonucun henüz belli olmamasından (örneğin bir ilacın yan etkisinin ne olacađın bilemeyebiliriz), veya sonucun daha önceden belli olmasına rađmen bunun farkında olmamamızdan kaynaklanmaktadır (örneğin, girdiđimiz bir sınavda başarılı olup olmadıđımızı bilemeyebiliriz). Olasılık, daha önceden bilinmeyen olayları arařtırmak için ortaya konmuř matematiksel bir arařtırma dilidir ve olasılık istatistik biliminin önemli bir parçasıdır.

Belirsizliğin çok belirgin olduđu tıp alanında istatistiğin tahmin fonksiyonu çok önemli bir yer tutmaktadır. Bir hastalık üzerine etki eden faktörlerin tahmin edilmesi, bir hastalıktan ölüme giden yolda etkili parametrelerin tahmin edilmesi veya bir hastalıktan iyileşmeye giden yolda etkili etmenlerin tahmin edilmesi tıbbi uygulamalar için inanılmaz bir öneme sahiptir. Bu amaçla kullanılan birçok istatistik yöntemi bulunmaktadır. Basit anlamda regresyon ve korelasyon analizinden ileri makine öğrenme yöntemlerine kadar bir çok yöntem tanımlanmıştır. Bu çalışmamızda bu amaca hizmet eden çok değişkenli regresyon ağaçlarından lojistik regresyon analizi (LRA), sınıflama ve regresyon ağacı (CART) ve rastgele orman yöntemleri tartışmaya çalışıldı ve örnek bir veri seti ile bu ağaçların performansları karşılaştırılmıştır.

2. GENEL BİLGİLER

2.1. Karar Ağaçları

Karar ağacı algoritması, veri madenciliği tekniklerinde önemli bir sınıflandırma yöntemidir. Bir karar ağacı, kök düğümü, dalları ve yaprak düğümleri olan bir ağaç gibi sınıflandırma ve regresyon modelleri oluşturur. Bağımlı değişken bir dikotomi olduğunda regresyon analizine alternatif bir yöntem olan lojistik regresyon, sınıflandırma amacıyla kullanılan bir diğer tekniktir.²

Karar ağacı, çok değişkenli verileri tanımlamak ve organize etmek için en yaygın kullanılan tekniklerden biridir. Bir karar ağacı genellikle bir veri madenciliği tekniği olarak kabul edilir. Güçlü yönlerinden biri, verileri diğer yöntemlerin yapamayacağı şekilde sınıflandırma yeteneğidir. Örneğin, doğrusal regresyon tarafından kaçırılacak doğrusal olmayan ilişkileri ortaya çıkarabilir. Bir karar ağacının anlaşılması ve açıklanması kolaydır, bu da bir araştırmacının konuya hâkim olmayan bir kitleye sonuçlarını bildirmesi gerektiği zaman büyük önem arz etmektedir. Karar ağaçları her zaman en iyi sonuçları üretmez, ancak iyi performans gösteren modeller ve basitçe açıklanabilecek modeller arasında makul bir uzlaşma sunarlar. Veri madenciliğinde kullanılan istatistiksel modeller Tablo 2.1.'de özetlenmiştir.³ Özellikle verilerin nasıl modelleneceğine dair çok az fikriniz olduğunda, karar ağaçları sadece modelleme için değil, aynı zamanda bir veri setini keşfetmek için de çok yararlıdır. Bir karar ağacı, bir veri kümesinin bir hedef değişkene (yani bağımlı bir y değişkenine) göre nasıl daha küçük gruplara ayrılacağını belirten hiyerarşik bir kural koleksiyonudur.

Bağımlı değişken kategorik ise, karar ağacına sınıflandırma ağacı denir. Bağımlı değişken sürekli ise, karar ağacına regresyon ağacı denir. Her iki durumda da ağaç tek bir kök düğümü ile başlar. Burada, düğümü bağımsız bazı değişkenlere göre bölmeye ve iki

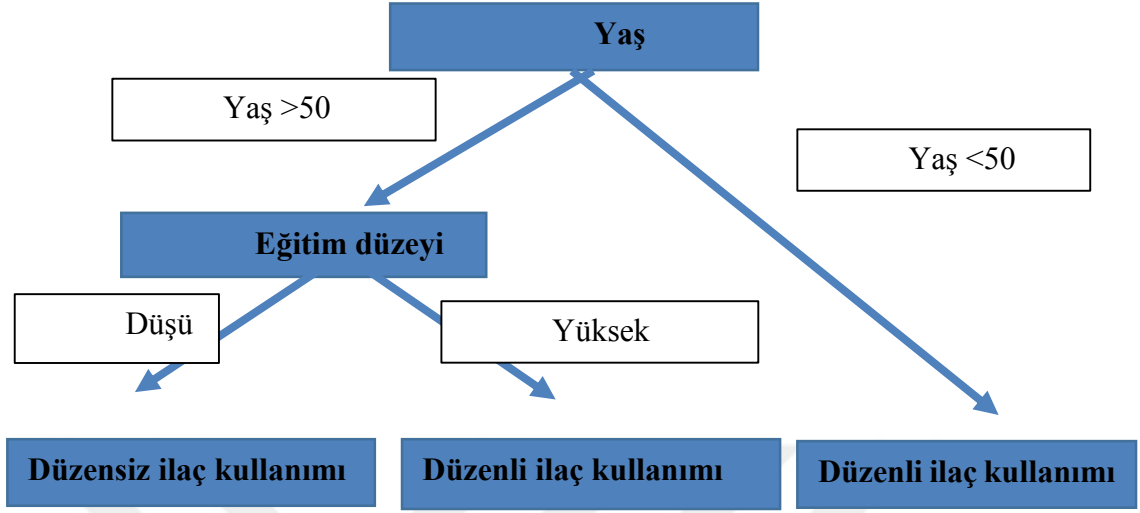
(veya daha fazla) yeni düğüm veya yaprak oluşturmaya karar verilir. Bu düğümlerin her biri, mümkünse benzer şekilde yeni düğümlere bölünebilir.⁴

Tablo 2.1. Veri Madenciliğinde Kullanılan Modeller

Veri Madenciliğinde Kullanılan Modeller			
Tahmin Edici Modeller		Tanımlayıcı Modeller	
Sınıflandırma	İstatistiksel Tahmin Modelleri	İlişki Analizi	Kümeleme Analizi
- Karar Ağaçları - Yapay Sinir Ağları - Genetik Algoritmalar - Bellek Tabanlı Sınıflandırma	- Regresyon Analizi - Diskriminant Analizi - Logistik Regresyon Analizi	- Birliktelik Kuralları - Ardışık Zamanlı Örüntüler	- Hiyerarşik Kümeleme - Hiyerarşik Olmayan Kümeleme

Hastaların ilaçlarını düzenli kullanma durumlarını ortaya koyan bir veri setimiz olduğunu varsayalım. Ve bunu öngörmek için bir karar verme algoritması oluşturmak istediğimizi düşünelim. Bunun için bir karar ağacı oluşturabiliriz. Şekil 2.1.'de gösterildiği gibi bir karar ağacı oluşturabiliriz. Bağımlı değişkenimiz ilaç kullanma durumu (düzenli veya düzensiz) kategorik olduğu için bu ağaç bir sınıflandırma ağacıdır. İlaç kullanma durumu için bağımsız değişkenlerimiz yaş ve eğitim düzeyi olsun. Önce yaşı iki gruba ayırırız: <50 yaş ve >50 yaş. Daha sonra eğitim düzeyin iki gruba ayırabiliriz (düşük eğitim düzeyi, yüksek eğitimi düzeyi). Geçmiş verilere göre, yaşı genç olan hastalar ilaçlarını daha düzenli kullandıkları tespit edilmiş. Yaşın genç olması eğitimden bağımsız olarak düzenli ilaç kullanımı için iyi bir gösterge olduğu anlaşıyor. Yaşı ileri olan hastalardan eğitim düzeyi yüksek olanların ilaçlarını daha düzenli kullandıkları görülmüş. Yaşı ileri ve eğitim düzeyi düşük olan hastaların ilaçlarını düzenli kullanmadıkları görülmüş. Bu şekilde elimizde yaş ve eğitim düzeyi bilgisi olan bir

hastanın ilaçlarını düzenli kullanıp kullanmadığına dair bir tahminde bulunabilir. Karar ağaçların çalışma mantığı bu şekildedir.⁵⁻⁶



Şekil 2.1. Basit bir karar ağacı örneği

Karar ağaçları, anlaşılma kolaylığı ve basit uygulama özellikleri gözönüne alındığında oldukça popüler makine öğrenme yöntemleri arasındadır.⁷

Karar analizinde, kararları ve karar vermeyi görsel ve açık bir şekilde temsil etmek için bir karar ağacı kullanılabilir. Karar ağacı, veri madenciliğinde kendisine yaygın bir kullanım alanı bulmuş bir yöntemdir.⁸ Amaç, birkaç bağımsız değişkene dayalı olarak bir bağımlı değişkenin değerini tahmin eden bir model oluşturmaktır.

Karar ağaçları veri madenciliğinde, bir veri setinin tanımlanmasını, sınıflandırılmasını sağlamak için kullanılan matematiksel yöntemlerin oluşturduğu bir model grubudur. Karar ağaçları, genel bir karar verme adımları uygulayarak, belirli sayıda veri barındıran bir veri kümesini çok küçük veri kümelerine ayırmaya yarayan bir modeldir.⁹ Her uygun ayırma işlemiyle, elde edilen grup bileşenleri diğerleriyle daha benzer duruma gelmektedir.

Bu yöntemde sınıflandırma amaçlı bir ağaç çizilir, sonrasında elimizdeki verilerdeki her bir kayıt, elde ettiğimiz ağaca uygulanır ve elde edilen sonuca göre

istediğimiz kayıt oluşturulur. Karar ağaçları yöntemi ile elimizde veri seti ne kadar karmaşık olursa olsun, bağımlı değişkenimize etki eden faktörleri ve bu faktörlerin modeldeki önemliliğini basit bir ağaç şeklinde görsel olarak sunabilme imkânı bulabilmekteyiz.^{10,11}

Karar ağacı tekniği ile verilerin sınıflanması genel olarak iki basamaktan oluşmaktadır.

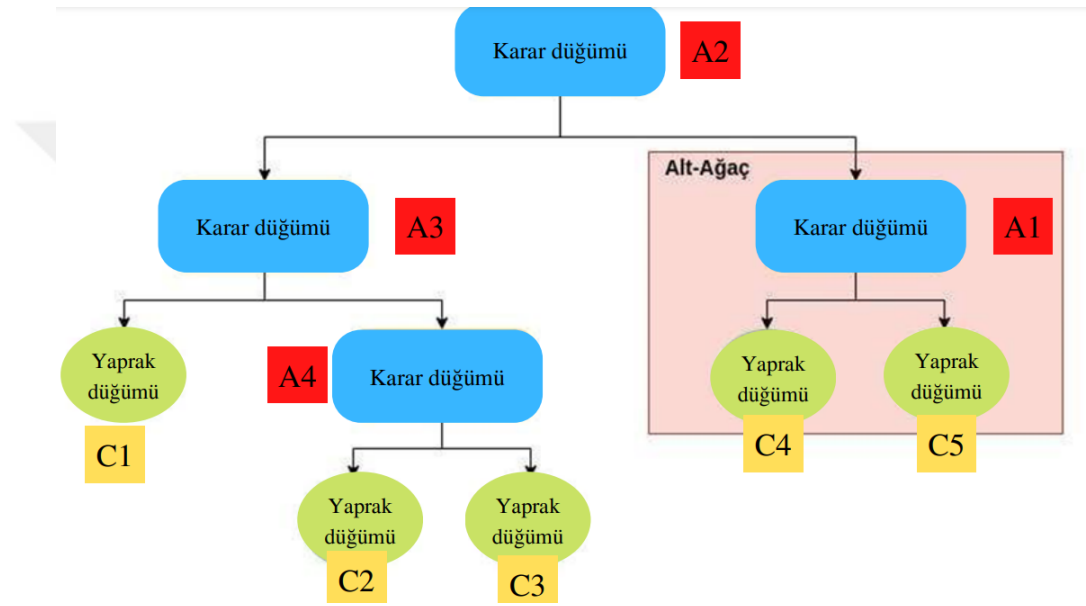
- Birinci basamak; öğrenme basamağıdır. Bu basamakta bir eğitim verisi kullanılarak model oluşturulur. Bu öğrenilen model ile sınıflama kuralları belirlenir.
- İkinci basamak ise daha önce oluşturduğumuz modelin uygunluğunu belirlemek için bu modelin test edilerek kullanıldığı sınıflamadır. Eğer modelin uygunluğu kabul edilebilir sınırlarda ise, eğitim modeli ile elde ettiğimiz kurallar daha sonraki verilerin sınıflaması için de kullanılabilir.¹⁰

Karar ağacı, anlaşıldığı üzere ağaç şeklinde bir ilerleme gösteren bir tekniktir. Ağacın yapısı; düğüm, dal ve yaprak denen bileşenlerden oluşur.¹²

Karar ağaçlarında yer alan düğümler Tablo 2.2.' de özetlenmiştir.³ Karar ağaçlarının yapısını aynen bildiğimiz ağaç yapısı gibidir. Kök, dallar ve yapraklardan oluşur. Tüm verileri kapsayan bir kök ile başlayarak normal ağacın tersine olarak kökü yukarıdan başlayarak yukarıdan aşağı doğru verileri daha küçük alt gruplara ayıran dallara ayrılırlar. Ağacın kökünden dallara doğru inen yapıda her boğum, “düğüm” (node) olarak adlandırılmaktadır, oluşan ağaçta, bu tarzda düğümler homojen değilse bu düğümlere “yavru düğümü”, eğer düğümler homojen ise “terminal düğüm” ismi verilmektedir.¹³

Tablo 2.2. Karar ağacı düğümleri

Düğüm ismi	Tanımlama	Ağaçtaki yeri
Kök düğüm	Ana düğümdür. Verideki bağımsız değişkenin bütün gözlemlerini kapsar	Başta
Ara düğüm	Karar düğüdür. Yaprak değildir. İlgili değişkenin aldığı değere göre tekrar bölünür. Bu nedenle uça bulunmaz	Ortada
Uç düğüm	Yaprak düğüdür. Bölünmez.	Sonda



Şekil 2.2. Karar ağacı yapısı

Yukarıdaki şekilde (Şekil 2.2.) karar ağacındaki $A1, A2, A3$ bileşenleri düğümleri oluşturmakta ve bu düğümler kendinden sonra iki ayrı dala ayrılmaktadır. Bu ayrılma esnasında, A_i düğümü ile ilgili bir soru sorulmakta (bu sonunun yanıtı veri tabanında olmalıdır) ve bu soruya verilen cevap ile de bir dal izlenmektedir. Ağaçtaki $C_1, C_2, \dots, C_m$ ile gösterilen düğümlerin her biri birer yaprağı temsil etmektedirler.

Karar ağaçları eğitici makine öğrenme yöntemleri içinde oldukça yaygın kullanılan bir tekniktir.

Bu tekniğin adımları:

1. Test öğrenme kümesinin oluşturulması,

2. Test kümesindeki örnekleri en iyi ayıran niteliğin belirlenmesi,
 3. Seçilen nitelik ile ağacın bir düğümünün oluşturulması ve bu düğümden çocuk düğümlerin ve yaprakların oluşturulması, çocuk düğümlere ait alt veri kümesinin örneklerinin belirlenmesi
 4. 3. adımda elde edilen her alt veri kümesi için
 - a. Örneklerin hepsi aynı sınıfa aitse;
 - b. Örnekleri bölecek nitelik kalmamışsa;
 - c. Kalan niteliklerin değerini taşıyan örnek yoksa; işlemi sonlandırılması
- gibi dört adımdan oluşmaktadır.¹³

2.1.1. Karar Ağacı Türleri

Veri madenciliğinde yaygın kullanılan karar ağaçlarının iki grubu vardır:

1. Sınıflandırma ağacı analizi, bağımlı değişkenler kategorik olduğu zaman kullanılır.
2. Regresyon ağacı analizi, bağımlı değişken sayısal bir değer olduğu zaman kullanılır. (Örneğin, bir evin fiyatı veya bir hastanın hastanede kalış süresi).

Sınıflandırma ve regresyon ağacı (CART) terimi, ilk olarak 1984 yılında Breiman ve arkadaşları tarafından tanıtılan yukarıdaki prosedürlerden herhangi birine atıfta bulunmak için kullanılan bir üst terimdir.¹⁴ Regresyon için kullanılan ağaçlar ve sınıflandırma için kullanılan ağaçların bazı benzerlikleri vardır. Fakat bölünmenin nerede olacağını belirlemek için kullanılan prosedürler birbirinden farklıdır.¹⁴

Genellikle topluluk yöntemleri olarak adlandırılan bazı teknikler, birden fazla karar ağacı oluşturur:

1. Güçlendirilmiş ağaçlar (boosted trees): Daha önce yanlış modellenen eğitim örneklerini vurgulamak için her yeni örneği eğiterek aşamalı olarak bir topluluk oluşturur. Tipik bir örnek AdaBoost'tur. Bunlar, regresyon tipi ve sınıflandırma

tipi problemler için kullanılabilir.^{15,16}

2. Önyükleme kümelenmiş (bootstrap aggregated) karar ağaçları, erken bir topluluk yöntemi, eğitim verilerini sürekli olarak yeniden örnekleyerek ve bir fikir birliği tahmini için ağaçları oylayarak birden çok karar ağacı oluşturur.¹⁷

- a. Rastgele orman(Rastgele bir orman sınıflandırıcısı), bu gruba ait bir alt sınıftır.

3. Rotasyon forest(Döndürme ormanı), her karar ağacının ilk önce girdi özelliklerinin rastgele bir alt kümesine temel bileşen analizi (PCA) uygulanarak eğitildiği bir yöntemdir.¹⁸

Karar ağaçlarında dallanma kriterlerine göre farklı bir sınıflandırma şekli de aşağıdaki gibidir.

1. Entropiye Dayalı Algoritmalar

- a. ID3 Algoritması
 - b. C4.5 Algoritması

2. Sınıflandırma ve Regresyon Ağaçları (CART)

- a. Twoing Algoritması
 - b. Gini Algoritması

Karar ağaçlarında düğüm ve dallanma kriterlerini belirleyen birçok özellik vardır. Ve bu özelliklere dayalı birçok algoritma geliştirilmiştir. Düğüm ve dallanma kriter seçimi için kullandıkları yol bakımından farklı özellikler olan bu algoritmalar aşağıda sıralanmıştır.^{19,20}

1. CHAID (Chi-Squared Automatic Interaction Detector: Otomatik Ki-Kare Etkileşim Belirleme),
2. CART (Sınıflama ve Regresyon Ağaçları),

3. MARS (Multivariate Adaptive Regression Splines: Çok Değişkenli Uyumlu Regresyon Uzanımları),
4. QUEST (Quick, Unbiased, Efficient Statistical Tree: Hızlı, Yansız, Etkin İstatistiksel Ağaç),
5. SLIQ (Supervised Learning in Quest),
6. SPRINT (Scalable Parallelizable Induction of Decision Trees)
7. ID3, C4.5 ve C5.0

2.1.2. Karar Ağaçlarının Üstünlükleri

Diğer veri madenciliği yöntemleri ile kıyaslandığında karar ağaçlarının kendine has çeşitli üstünlüklerinden bahsetmek mümkündür;

1. Anlaşılması ve yorumlanması kolaydır. Okuyucular kısa bir açıklama ile karar ağacı modellerini kolayca anlayabilirler. Ağaçlar ayrıca, ilgili konuda uzman olmayanların kolayca yorumlayabileceği bir şekilde görsel olarak da gösterilebilir.²¹
2. Hem sayısal hem de kategorik verilerin işlenmesi mümkündür.²¹ Diğer teknikler genellikle yalnızca tek çeşit değişkene sahip veri kümelerinin analizinde kullanılabilir. (bazı yöntemler sadece sayısal değişkenleri yorumlayabilirken, bazıları sadece kategorik değişkenleri analiz edebilir)
3. Veri hazırlığı için çok zaman ve efor gerektirmez. Diğer tekniklerde genellikle veriler için normal dağılım şartı gerekir . Ağaçlar nitel tahmin edicileri işleyebildiğinden, kukla değişkenler oluşturmaya gerek yoktur.²¹
4. İstatistiksel testler kullanarak bir modeli doğrulamak mümkündür. Bu, modelin güvenilirliğini ortaya koymayı mümkün kılar.
5. Büyük veri kümeleriyle iyi performans gösterir. Standart bilgi işlem kaynakları kullanılarak makul bir sürede büyük miktarda veri analiz edilebilir.

6. İnsan karar verme sürecini diğer yaklaşımlara göre daha iyi yansıtır.²¹Bu, insan kararlarını/davranışlarını modellerken faydalı olabilir.

2.1.3. Sınırlılıkları

1. Ağaçlar çok sağlam olmayabilir. Eğitim verilerindeki küçük bir değişiklik, ağaçta ve dolayısıyla nihai tahminlerde büyük bir değişikliğe neden olabilir.²¹
2. Karar ağacı öğrenenler, eğitim verilerinden iyi genellemeyen aşırı karmaşık ağaçlar oluşturabilirler. Bu, fazla uydurma olarak bilinir.²² Bu sorundan kaçınmak için budama gibi mekanizmalar gereklidir ^{19,20}
3. Sınıflandırmaya kadar düğüm sayısı veya testler tarafından tanımlanan ağacın ortalama derinliğinin, çeşitli bölme kriterleri altında minimum veya küçük olması garanti edilmez.²³
4. Farklı düzey sayılarına sahip kategorik değişkenleri içeren veriler için, karar ağaçlarındaki bilgi kazanımı, daha fazla düzey içeren nitelikler lehine önyargılıdır.²⁴

Son yıllarda ilişkilerin birçok yerde doğrusal olmaması ve incelenen değişkenlerin çoğunun normal dağılmaması nedeniyle doğrusal olmayan, hiyerarşik, kural tabanlı yöntemlerin kullanımında artış olmuştur. Veri madenciliğinde yer alan sınıflandırma ve karar ağaçlarının amacı, verilere dayalı bir model geliştirmek ve veri setlerinin sonuç değerlerini tahmin etmektir.²⁵ İstatistiksel analiz yöntemlerinde verinin yapısına göre uygulanacak en uygun yöntemin belirlenmesi çok önemlidir. Dolayısıyla lojistik regresyon ve karar ağacı sınıflandırmasının günümüzde kullanılan temel sınıflandırma algoritmaları olduğu söylenebilir. Bazı çalışmalar, sınıflandırma ve regresyon ağaçlarının literatürdeki standart yaklaşımdan daha iyi performans gösterdiğini, bazıları ise karar ağaçlarının lojistik regresyondan daha iyi bulunduğunu göstermektedir. Hiçbiri diğerinden daha iyi değildir ve her birinin performansı genellikle verilerin doğasına bağlı

olarak deęiřir. Bu yöntemlerden en çok tercih edileni sınıflandırma ve regresyon ağacı tekniğidir. Sınıflandırma ve regresyon ağacı analizi, bir tür karar ağacı metodolojisidir. Sınıflandırma ve regresyon ağaçlarının açıklanması kolaydır ve model sonuçlarının yorumlanması kolay bir görselleřtirmesini sağlar.²

2.2. Lojistik Regresyon Analizi

Lojistik regresyon analizi, son dönemlerde kullanımı sıklıklařan bir regresyon yöntemidir. Neden sonuç ilişkilerini açığa çıkarmayı amaçlayan tıbbi arařtırmalarda deęiřkenler evet hayır, hasta-saęlam, öldü-ölmedi, başarılı- başarısız, var-yok gibi dikotom verilerden oluřmaktadır. Lojistik regresyon analizi, baęımlı deęiřkenin dikotomi-sıralı-kategorik olduęu ve belirli bir daęılım varsayımı beklemeksizin baęımlı deęiřken ile baęımsız deęiřkenler arasındaki neden sonuç ilişkisinin ortaya konmasında ve aynı zamanda baęımsız deęiřkenlerin etkilerine dayanarak verilerin sınıflandırılması yarayan bir yöntemdir.²⁶

Lojistik regresyon, baęımlı ve baęımsız deęiřkenler arasında bir logit ilişkisi olduęunu varsayar. Lojistik Regresyon, verilerin uzayda (kıvrımlı) doğrusal olarak ayrılabilir olduęunu varsayar. Bu nedenle, lojistik regresyon, doğrusal olmayan (üstel veya polinom) modeller üretebilir. lineer regresyonda baęımlı deęiřken tahmini bir deęere sahipken, lojistik regresyonda baęımlı deęiřkenin alabileceęi deęerlerden birinin gerçekteřme olasılıęı tahmin edilmektedir. Bu yöntemin en büyük avantajı olasılıklı bir sınıflandırma formülü üretmesidir.^{27,28} Lojistik regresyon modeli ařaęıdaki formül ile gösterilmiřtir.

$$L = \ln(pi/(1 - pi)) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k$$

Burada 0 sabit deęerdir ve bi (i = 1, 2,... k) her baęımsız deęiřkenin regresyon katsayılarını gösterir.

$\ln(pi/(1 - pi))$ lojistik dönüşümü temsil eder ve pi arzu edilen durumun gerçekleşme olasılığını gösterir. Böylece $pi/(1 - pi)$ odds oranı olarak adlandırılır. Olasılık değeri pi aşağıdaki formül ile gösterilir.

$$p_i = \exp(\sum_{k=0}^n \beta_k x_{ik}) / (1 + \exp(\sum_{k=0}^n \beta_k x_{ik}))$$

Regresyon analizinde bağımlı değişken kategorik veya sürekli olabilir. Bağımlı değişken sürekli ise analiz lineer regresyon, bağımlı değişken kategorik ise analiz lojistik regresyon olarak adlandırılır.^{29,30} İki grup varsa, ikili lojistik regresyon kullanılır.²

Lojistik regresyon analizi biyomedikal araştırmalarda çok sık kullanılan popüler bir matematiksel modelleme yöntemidir. Hangi regresyonun seçileceği bağımlı değişkene bağlı olduğundan dolayı, bağımlı değişkenin ikili (binary, dichotomous) olduğu durumlarda en fazla tercih edilen yöntem lojistik regresyondur.³¹

Lojistik regresyon; bağımlı değişkeninin kategorik olduğu durumlarda, bağımlı değişkenler ile bağımsız değişkenler arasındaki neden sonuç ilişkisini ortaya koymak için kullanılan bir yöntemdir. Bu kategoriler ikili, üçlü veya daha fazla sayıda olabilir. Bağımlı değişkeninin beklenen değerlerinin bağımsız değişkenlere göre olasılık olarak hesaplandığı bir regresyon yöntemidir. Bağımlı değişken ile bağımsız değişken arasındaki matematiksel ilişkiyi analiz etmek için basit ve çoklu regresyon analizleri kullanılmaktadır. Bu yöntemlerin uygulanabilmesi için veri setlerinin sağlaması gereken birtakım varsayımlar vardır.

1. Bağımlı değişkenin normal dağılım göstermesi,
2. Bağımsız değişkenlerin normal dağılım gösteren örneklemelerden seçilmiş olması
3. Hata varyansının normal dağılım göstermesi gerekmektedir.

Bu varsayımların sağlanamadığı durumlarda basit ya da çoklu regresyon analizlerinin kullanılması uygun değildir. Lojistik regresyon analizi, sınıflama ve atama

işlemi imkânı sunan bir regresyon yöntemidir. Varsayımları içinde normal dağılım, süreklilik ön şartı yoktur. Bağımsız değişkenlerin bağımlı değişken üzerinde etkileri olasılık olarak elde edilir ve sonrasında risk faktörlerinin olasılık olarak belirlenmesine çalışılır.³²⁻³⁵ Araştırmacılar çalıştıkları değişken etkili birden çok etken olduğu durumlarda bu etkenlerin bağımlı değişken üzerine tek tek etkisinin yanı sıra, bunların hep birlikte bağımlı değişken üzerine olan etkisini de öğrenmek ve analiz etmek istemektedirler. Bu birlikte etkinin incelenmesi için elimizde değişik istatistik yöntemler bulunmaktadır.

Örneğin, bağımlı değişkenin sürekli olduğu, bağımsız değişkenlerin kesikli olduğu durumlarda varyans analizi, bütün değişkenlerin kesikli olması durumunda log-linear modeller, bütün değişkenlerin sürekli olduğu durumlarda regresyon analizleri gibi yöntemler bulunmaktadır.

Sağlık bilimlerinde yapılan çalışmalarda değişkenlerin tür ve yapısı biraz değişkenlik gösterebilir, sürekli ve kesikli karışımı bağımsız değişkenlerle karşılaşmak mümkündür. Tıp alanında en çok üzerinde durulan ve çoğu araştırmanın asıl amacını oluşturan şey etkenlerle hastalık arasındaki ilişkinin risk bakımından irdelenmesidir. Bu durumlarda kullanılan yöntem LRA'dır.^{36,37}

LRA'nın tarihsel gelişimine baktığımız zaman, lojistik regresyon modelinin kullanımı 1944 yılında Berkson ile başlamıştır ve 1972 yılında Finney tarafından probit analize alternatif bir seçenek olarak sunulmuştur.

Lineer regresyon analizinde amaç bağımlı değişkenin alacağı değerleri tahmin etmek iken, LRA'da amaç bağımlı değişkenin alabileceği değerlerin gerçekleşme olasılığını tahmin etmektir.

Bu değeri tahmin etmek için kullanılan model aşağıda verilmiştir.³⁸

$$\pi(\chi) = \frac{e^{\beta_0 + \beta_1 \chi}}{1 + e^{\beta_0 + \beta_1 \chi}}$$

Lojistik regresyon modeliyle tahmin yapabilmek için kullanılan yöntem genel olarak en çok olabilirlik metodudur. Basit anlamda en çok olabilirlik metodu, gözlenen veri setini elde etmenin olasılığını en yüksek yapan bilinmeyen değerleri tahmin etmede kullanılır. Bu metodun uygulanabilmesi için fonksiyonun oluşturulması gerekir. Aşağıdaki ifade LRA'da tahmin edilen değerler ile gözlenen değerlerin karşılaştırılması için kullanılır.³⁸

$$D = -2 \ln \frac{\text{Şuandaki Modelin Olabilirliği}}{\text{Doymuş Modelin Olabilirliği}}$$

Modeldeki bir bağımsız değişkenin önemliliğini belirlemek için formülde o bağımsız değişkenin var olduğu ve olmadığı durumlardaki D değerleri, G istatistiği ile karşılaştırılırlar. G istatistiği ki-kare dağılımı gösterir ve serbestlik derecesi p dir.^{38,39}

$$G = D(\text{Değişkensiz Model için}) - D(\text{Değişkenli Model için})$$

$$G = D(\text{Değişkensiz Model için}) - D(\text{Değişkenli Model için})$$

$$D = -2 \ln \frac{\text{Değişkensiz Modelin Olabilirliği}}{\text{Değişkenli Modelin Olabilirliği}}$$

2.3. Sınıflandırma ve Regresyon Ağacı (CART)

Karar ağaçları, problem oluşturan veri kümelerinin yapısına göre sorular sorarak ve karar kuralları oluşturarak bir ağaç yapısı gibi sınıflandırma ve regresyon modelleri oluşturur. Bu işlem için ağaç yapısının temel elemanı olan kök düğümde sorular sorulmaya başlanır ve ağaç yapısının son elemanı olan yapraklara ulaşana kadar ağaç dallanarak büyür.^{40,41} Ağaç yapısı kök düğümün en önemli bağımsız değişken olduğunu, alt dallar ise diğer bağımsız değişkenler olduğunu göstermektedir. Böylece yapraklar

bağımlı değişkenin değerlerini gösterir. Karar ağaçları içerisinde yer alan sınıflandırma ve regresyon ağaçları, kolay anlaşılması, varsayım gerektirmemesi ve sonuçların diyagramlar halinde görsel olarak sunulması gibi avantajları nedeniyle sıklıkla tercih edilmektedir.⁴² İncelenen veri seti üzerinde herhangi bir varsayım gerektirmediği için parametrik tekniklere güçlü bir alternatiftir. Tüm bağımsız değişkenler kullanılarak veri setinin alt kümelerine bölünmesiyle oluşturulan sınıflandırma ve regresyon ağacı, hem regresyon hem de sınıflandırma için kullanılacak özyinelemeli bir bölümlenme yöntemidir. En iyi tahmin edici değişkeni seçmek için birtakım ölçüler mevcuttur. Bu ölçüler gini, en küçük kareler sapma, entropi, twoing, bilgi kazancı ve bilgi kazanç oranı, ki kare olasılık değeri gibi ölçülerdir.²⁸ Buradaki amaç, değişkenler için mümkün olan en homojen veri alt gruplarını üretmektir. Sınıflandırma ve regresyon ağaçları homojen alt sınıflar için bağımlı değişkenler ve bağımsız değişkenler arasındaki ilişkiyi veri setindeki heterojenliği dikkate alarak modelleyebilir ve bu ilişkiyi bir ağaç yapısı şeklinde görselleştirebilir. Sınıflandırma ve regresyon ağaçları ile hem kategorik hem de sürekli değişkenler modellenebilir. Bağımlı değişken kategorik veya sürekli olabilir. Eğer bağımlı değişken kategorik ise analize sınıflandırma ağacı, bağımlı değişken sürekli ise analize regresyon ağacı denir.⁴³ Regresyon ağaçlarında sınıf yoktur ve veriler sürekliktir. Bu nedenle regresyon ağacı tekniğinde sınıflandırma, ayırma kuralları, gini indeksi veya entropi ölçümü uygulanamaz. Regresyon ağacındaki dallanma işlemi, artıkların karelerini azaltma algoritmasına göre gerçekleştirilir.⁴²

CART'ın ilk uygulamaları 1984 yılında Breiman ve arkadaşları tarafından yapılmış ve bir karar ağacı algoritması olarak bilimsel alanda kullanıma sunulmuştur.⁴⁴

CART algoritması her adımda ilgili grubun, kendisinden daha homojen iki alt gruba ayrılmasını sağlamaktadır. Yani her bir dal ikili alt gruplara ayrılarak büyüme gerçekleşir. Ayırma işleminde bağımlı değişken kategorik ise gini veya twoing, bağımlı

değişken sürekli ise en küçük kareler sapmasından yararlanılmaktadır. Tüm veri tipleri için uygu olan CART algoritmasında asıl nokta, karar noktalarında birimlerin homojen sınıfları oluşturacak şekilde dallanmasıdır.⁴⁵

CART algoritması her değişken tipine uygun olduğu için normallik, dönüşüm varsayımları sağlama zorunluluğu yoktur. CART algoritması kayıp veriler için o verilerin yerine temsili bir değer atayabilmektedir. Sınıflandırma ve regresyon amaçlı kullanılan CART algoritması entropi temellidir ve dallanma kriteri için kullanılan teknikler: Twoing ve Gini indeksleridir.⁴⁶

2.3.1. Gini Ayırma Kriteri

k : Sınıfların _ sayısı

T : Bir _ düğümdeki _ kayıtların _ sayısı

T_{sol} : Sol _ taraftaki _ kayıtların _ sayısı

$T_{sağ}$: Sağ _ taraftaki _ kayıtların _ sayısı

L_i : Sol _ tarafta _ i _ kategori sin deki _ kayıtların _ sayısı

R_i : Sağ _ tarafta _ i _ kategori sin deki _ kayıtların _ sayısı

$$Gini_{sağ} = 1 - \sum_{i=1}^k \left(\frac{L_i}{T_{sağ}} \right)^2$$

$$Gini_{sol} = 1 - \sum_{i=1}^k \left(\frac{R_i}{T_{sol}} \right)^2$$

Nitelik değerlerini iki parçaya ayırarak dallanma yapan gini ölçeği, bölünme noktalarında $Ginisağ$ ve $Ginisol$ değerlerini hesaplamaktadır. Her dallanmadan sonra da Gini değeri en küçük olan seçilmektedir. Bu işlemin amacı, her adımda en büyük veri kümesini oluşturmaktır. Bu şekilde en iyi dallanma sağlanmış olmaktadır.⁴⁷

Her j niteliği için, veri setindeki satır sayısı n olmak üzere Gini endeks değeri formülle hesaplanmaktadır.

$$Gini_j = \frac{1}{n} - T_{sol} \times Gini_{sol} + T_{sağ} \times Gini_{sağ}$$

2.3.2. Twoing Ayırma Kriteri

Twoing ayırma kriteri gini indeksinden farklı olarak tek bir sınıfı diğerinden ayırmak yerine, düğüme ait verilerin yarısını barındıran ve birbirine göre daha heterojen sınıflar oluşturan bir dallanma imkânı sunmaktadır. Bundan dolayı, Twoing kriteri Gini kriterine göre daha dengeli bir dallanma yaparken, dolayısıyla işlem süresi de uzamaktadır. Bir t düğümünde sol $tsol$ ve sağ $tsağ$ şeklinde iki dal bulunur. Regresyon ağacı oluşturulmasında kullanılacak her bir veri sağ ve sol dala bölünmeye adaydır. Twoing kriterinde her bir aday için sağ ve sol taraftaki dalda bulunma olasılıklarının her biri için hesaplamalar aşağıdaki gibidir.⁴⁸

t : Dallanmanın _yapılacağı_ düğüm

P_{sol} : Veri _setindeki_ kaydın _solda_ olma _olasılığı

$P_{sağ}$: Veri _setindeki_ kaydın _sağda_ olma _olasılığı

$P\{jvt_{sol}\}$: j _sınıf_ _değerinin_ _sol_ _taraftaki_ _bölünmede_ _olma_ _olasılığı

$P\{jvt_{sağ}\}$: j _sınıf_ _değerinin_ _sağ_ _taraftaki_ _bölünmede_ _olma_ _olasılığı

Yukarıda değerlere göre t düğümündeki s aday bölünmelerinin uygunluk ölçüsü aşağıda gösterilmektedir.

$$\phi(st) = 2P_{sol}P_{sağ} \sum_{j=1}^n \left[P(jIt_{sol}) - (jIt_{sağ}) \right]$$

Bu işlem sonrasında elde edilen en büyük uygunluk ölçüsü seçilmektedir. Bu ölçüye göre kök düğümde elde edilecek niteliğin hangisi olacağına karar verilmekte ve bu şekilde dallanma işlemi yapılmaktadır. Yukarıdan aşağıya doğru ağacın tüm yapraklarına ininceye kadar aynı işlem tekrar edilmektedir.⁴⁹

2.3.3. En Küçük Kareler Sapması (LSD)

CART yönteminde bağımlı değişken sürekli sayısal bir değişken olduğunda yani regresyon ağacı oluşumunda kullanılan ve safsızlık fonksiyonuna dayalı bölme kriteri en küçük kareli sapma yöntemidir. Sürekli bağımlı değişkenler için kullanılan en küçük kareli sapma heterojenlik ölçüsüdür. Bu ölçü $R(t)$, t düğümü için basit (ağırlıklandırılmış) düğüm içi varyansdır ve düğüm için risk tahminine eşit olmaktadır.⁴⁹

$$R(t) = \frac{1}{N_w(t)} \sum_{i \in t} W_n f_n \left[y_1 - \{y'(t)\}^2 \right]$$

$N_w(t)$: t düğümündeki ağırlıklandırılmış durum sayısı

W_n : Ağırlıklandırılmış değişken değeri

f_n : frekans değişkeninin değeri

y_1 : hedef değişken değeri

$y'(t)$: t düğümü için ağırlıklı ortalama

CART yöntemi ile ağaç oluşturmak için üç basamak vardır. Bunlar: “örnekleme bölme kuralı”, “bölme uyum kriteri” ve “optimum ağacı seçme” kriteridir. CART ile ağaç oluşturabilmek için önceden belli bir bölme kuralı mevcuttur. Bu kural düğüm bölme sürecinin her aşamasında uygulanır. Önceden belirlenmiş bölme kuralına ve düğüm bölme sürecinin her aşamasında uygulanan bölme uyum kriterine göre ağaç oluşturur. CART ile ağaç oluştururken başlangıç veri seti, bağımsız değişkene sorulan evet/hayır yanıtlarına göre alt dallara bölünerek ağaç oluşturulur. Her bir bağımsız değişkene soru sorulur. Bağımsız değişkenden alınan yanıtla göre veriler sağ ve sol iki alt dala ayrılır.^{50,51} CART algoritması bütün bağımsız değişkenlerin mümkün olabilecek bütün değerlerini araştırarak en optimum bölünmeyi bulmaya çalışır. Başlangıç veri setine sorulan sorulara verilen yanıtlar ile veri seti daha küçük alt birimleri ayrılır. Soru sormanın amacı veriyi

en yüksek seviyede homojenize olmuş iki alt birime bölmektir. Bu anlatılan işlemler oluşan her bir alt düğüm için tekrarlanır ve uç düğüm oluşana kadar devam eder.⁵² Bu şekilde CART algoritması ile veri setini en düşük hata seviyesi ile sınıflama yapan ideal ağaç elde edilmiş olur. Özetle CART algoritması bütün başta yer alan veri setini içeren kök düğümden itibaren her düğümü iki çocuk düğüme böler ve bu şekilde ikiye bölünmüş ağaçlar oluşturur. CART algoritmasının temel amacı homojenliği maksimum seviye çıkarmaktır. Bir düğümün içinde eğer homojen bir alt birim varsa o düğümün safsızlığına işaret eder. Başka bir deyişle bir uç düğüm her koşulda aynı değeri alıyorsa bölünme yapmaz.⁵³ Safsızlığın ölçüsü değişkenin tipine göre değişkenlik gösterir. Kategorik bağımsız değişkenler için safsızlık ölçüsü Gini ve Twoing kriterleridir. Sürekli bağımsız değişkenler safsızlık ölçüsü olarak en küçük kareli sapma yöntemi kullanılır. Bu kriterler safsızlık fonksiyonuna bağlı bölünme kriterleridir.

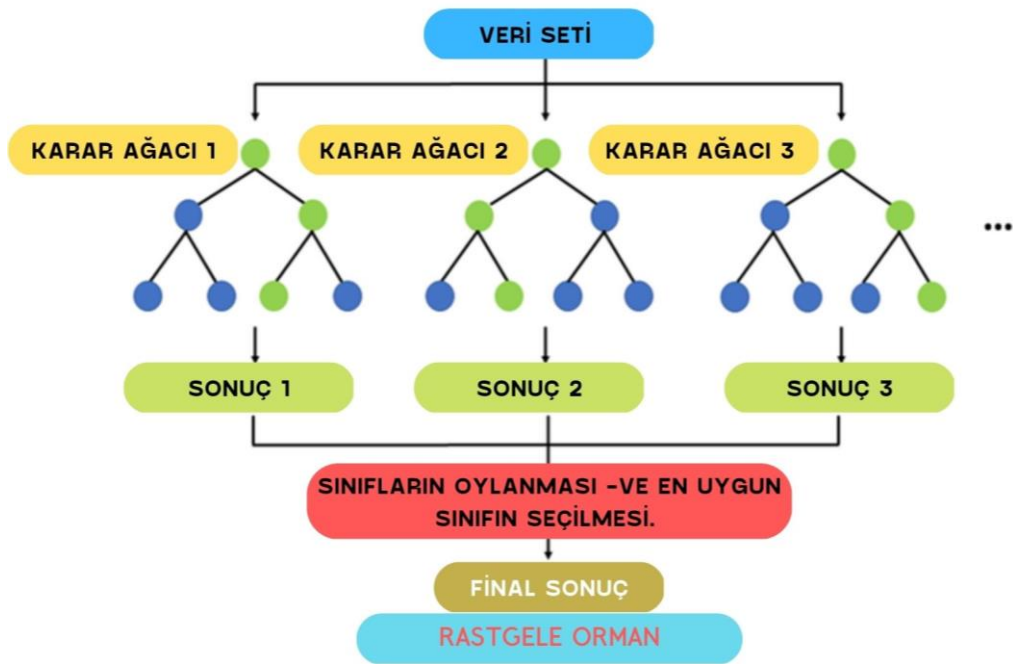
2.4. Rastgele Orman Algoritması

Rastgele Orman Algoritması, eğitimde ve testlerde çok sayıda karar ağacı oluşturarak, sınıfların modu olan sınıfın çıktısını aldığı bir sınıflandırma algoritmasıdır. Karar ağacı, veri özelliklerinden çıkarılan basit karar kurallarını öğrenerek çalışır. Ağaç ne kadar derin olursa, karar kuralları o kadar karmaşık ve model o kadar uygun olur. Rastgele karar ormanları, karar ağaçlarının aşırı uydurma sorununun üstesinden gelir. Bir girdi vektöründen yeni bir kaydı sınıflandırmak için girdi vektörünü ormandaki ağaçların her birine koyun. Her ağaç belirli bir sınıf etiketine oy verir. Son sınıf etiketi, ormandaki tüm ağaçlar üzerinde en çok oyu alan etiket olacaktır. Uygulamada, rastgele orman oldukça hızlı bir sınıflandırma tekniğidir ve dengesiz büyük veri kümeleriyle başa çıkabilir.⁵⁴

Sınıflandırma görevleri için, rastgele ormanın çıktısı, çoğu ağaç tarafından seçilen sınıftır. Regresyon görevleri için, tek tek ağaçların ortalama veya ortalama tahmini

kullanılır.^{55,56} Rastgele karar ormanları, karar ağaçlarının eğitim setlerine fazla uyum problemlerini düzeltir.⁵⁷ Rastgele ormanlar genellikle karar ağaçlarından daha iyi performans gösterir, ancak doğrulukları eğitim destekli ağaçlardan daha düşüktür]. Ancak, veri özellikleri performanslarını etkileyebilir.^{58,59}

Rastgele karar ormanları için ilk algoritma, 1995 yılında Tin Kam Ho oluşturuldu.^{55,60-62} Algoritmanın bir uzantısı Leo Breiman ve "Rastgele orman"ı 2006 yılında bir ticari bir marka olarak kaydeden Adele Cutler tarafından geliştirilmiştir.⁶³⁻⁶⁶ Rastgele ormanlar, çok az yapılandırma gerektirirken geniş bir veri yelpazesinde makul tahminler ürettikleri için işletmelerde sıklıkla "kara kutu" modelleri olarak kullanılır



Şekil 2.3. Rastgele orman akış şeması

Rastgele orman, torbalama (bagging) temelli bir kolektif (ensemble) karar ağacı yöntemidir.⁶⁷ Rastgele orman, diğer karar ağacı yöntemleri gibi tüm bağımsız değişkenler arasından en uygun dalı kullanarak her bir düğümü dallara ayırmaktan ziyade, mevcut olan her bir düğümde random(rastgele) olarak seçilen değişkenler içinden en uygununu

seçerek her bir düğümü dallara ayırır. Oluşan veri setleri orijinal veri setinden tekrardan örnekleme kullanılarak oluşturulur.⁶⁸ Basitçe anlatmak gerekirse: rastgele orman yönteminde, birden fazla karar ağacı oluşturulur ve bu ağaçlar daha doğru bir tahminde bulunmak için birleştirilir.

2.4.1. Rastgele Orman Algoritması Akış Şeması

Rastgele orman algoritması için araştırmacının iki parametre tanımlaması gerekir.

Bu parametreler:

1. Geliştirilecek ağaç sayısı (N)
2. Her bir düğümde kullanılacak değişken sayıdır (m).

Rastgele orman karar ağacı yönteminde ağaç dallanmaları kullanılacak algoritma ID3 algoritmasıdır.

ID3 algoritması rastgele orman için oluşturma için kullanılır ⁶⁷ ve belirsizlik ölçümü kriteri için ID3 algoritması kullanılır ve aşağıdaki şekilde formüle edilir.

$$H(T) = -\sum_{i=1}^n p(x_i) \log p(x_i)$$

H entropi için kullanılır. Entropi belirsizliğin ölçüsüdür ve bu değer ne kadar yüksekse, belirsizlik de o kadar yüksektir. Entropi = 0 oluncaya kadar dallanma devam eder. Şekil 2.3. de klasik bir rastgele orman algoritması gösterilmiştir.

ID3 algoritması ile ağaç oluştururken iki kavramdan yararlanır. Entropi ve bilgi kazancı kavramları.

1. Entropi: Belirsizliğin ölçüsüdür; bu ölçü ne kadar yüksekse, belirsizlik o kadar fazladır. Formülü aşağıdaki gibidir.

$$H(T) = -\sum_{i=1}^n p(x_i) \log p(x_i)$$

2. Bilgi Kazancı: Hangi özelliğin bölünme oluşturmak için entropiyi en iyi şekilde kullanacağını belirlemek amacıyla kullanır. Formülü;

$$Kazanç(T, A)=H(T)-H(T,A)$$

Bilgi kazancı, dallanma olmadan önceki veri setinin entropisi ile her bir özneliğin entropisi arasındaki fark olarak tanımlanabilir.

ID3 algoritması aşağıdaki gibidir.⁶⁹

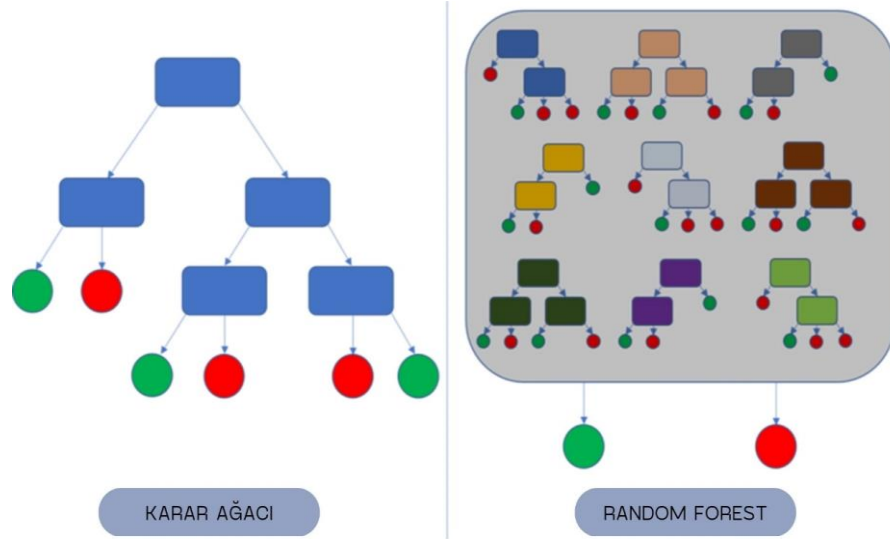
Önce kök düğüm oluşturulur. Bu kök düğüm tüm öğrenme kümesini alt kümesi olarak içermelidir.

1. Bilgi kazancını hesaplanmalıdır. (her bir öznelik için)
2. Dallanma için maksimum bilgi kazancı ile özneliği seçilmesi gerekir.
3. Entropi = 0 olan yaprak düğüme ulaşılan kadar bu kökün her alt düğüme

ID3 algoritmasını uygulanarak ağacın oluşturulması gerekir.

Rastgele orman algoritması basamakları aşağıdaki gibidir.⁷⁰

- Orijinal veri setinden N adet örnekleme seçilir.
- Seçilen örneklemin belli bir kısmı oluşturulacak ağaçlar için öğrenme verisi olarak kullanılır
- Budanmamış sınıflama ağacı aşağıdaki adımları takip ederek oluşturulur.
 - in bag veri setinden her düğümde bütün öznelilikler içerisinde en iyi özneliği seçmek yerine rastgele m tane öznelik seçilir ve bunların içerisinde en iyi dallara ayıracak (en çok bilgi kazancı sağlayacak) olanı belirlenir.



Şekil 2.4. Karar ağacı ve rastgele orman karşılaştırılması

2.5. Stres Ekokardiyografi

Stres ekokardiyografi koroner arter hastalığının tanısında, mevcut bir iskemik hastalığın değerlendirilmesinde, sol ventrikül miyokardın canlılık araştırılmasında ve kardiyak dışı cerrahi öncesinde risk değerlendirilmesinde önemli bir yer tutmaktadır. Klasik egzersiz en çok kullanılan stres indükleyicisi olarak kullanılmaktadır. Fakat egzersiz yapamayanlarda dopamin, dobutamin gibi ilaçlar ve atriyal pacing girişimler de stres indükleyici olarak kullanılabilir. ^{71,72} Stres ekokardiyografisinin yukarıdan belirtilen amaçlar için kullanılan diğer yöntemlerden en önemli avantajları diğerlerinde daha ucuz ve tekrarlanabilir olması, radyasyon gibi bir dezavantajının olmamasıdır. ⁷³ Standart egzersiz stres testinin kısıtlılıklarını aşmak için uygun bir görüntüleme çözümü ararken, stres ekokardiyografi pratik açıdan cazip bir yöntemdir. Kalbin invaziv olmayan görüntülenmesini sağlayan yaygın ve ucuz bir tekniktir. Stres ekokardiyografisinde çeşitli stresörler kullanılarak stres kaynaklı duvar hareketi anormalliklerinin saptanması yoluyla miyokard iskemisini tanımlamanın bir yolunu sağlar. Önemli koroner darlıkların tespiti için doğruluk oranı %80-90 arasında değişmektedir. Stres ekokardiyografi, kronik koroner hastalıkta, miyokard enfarktüsünden sonra ve majör kalp dışı cerrahi öncesi

hastaların değerlendirilmesinde güçlü bir prognostik araçtır. Revaskülarizasyondan sonra dissinerjik bölgelerin fonksiyonel iyileşmesinin öngörülmesi için uygun bir testtir ve ayrıca kapak cerrahisi düşünülen hastalarda değerli fizyolojik bilgiler sağlar.

Bu tekniğinde birtakım dezavantajları mevcuttur. Fakat görüntüleme ve görüntü işlemedeki ilerlemeler, görüntü kalitesiyle ilgili sorunların hepsini olmasa da çoğunu çözmüştür.

Test yorumu, değerlendiren kişiye göre değişebilmesi bir dezavantajdır. Testin sonucu için tek ana belirleyici iskemiye bağlı anormal duvar hareketi olduğu için metabolik anlamda iskemi indüklemeye ihtiyacı, stres ekokardiyografinin maksimum düzeyde egzersiz yapan veya antianjinal tedavi gören hastalarda koroner arter hastalığını saptamadaki doğruluğunu sınırlar.

Ultrason ve dijital teknolojideki teknolojik gelişmelerin tekniği daha da geliştirmesi ve basit bir duvar hareketi testinden lokal kasılma değerlendirmesi ve perfüzyonu değerlendirme kabiliyeti kazanmasına yol açabilir.

2.5.1. Stres Ekokardiyografi Patofizyoloji

Egzersiz ve inotropik stres normalde kalbine bölgesel duvar hareketinde ve kalınlaşmada genel bir artışa neden olur ve ejeksiyon fraksiyonunda bir artış izlenir. Bölgesel sistolik duvar hareket bozukluğu genellikle koroner arter hastalığını düşündürür fakat kardiyomiyopatiler de duvar hareketinde bölgesel farklılıklara yol açabilir.

İstirahat sol ventrikül duvar hareket kusurları geçirilmiş miyokard enfarktüsünün ayırt edici özelliğidir, ancak mutlaka bu segmentin canlı olmadığı anlamına gelmez.

İskemi tipik olarak yeni veya kötüleşen duvar hareket anormallikleri, gecikmiş kasılma veya sol ventrikül genişlemesinin gelişmesi veya ejeksiyon fraksiyonunda azalma olarak kendini gösterir. Miyokard infarktüsü esnasında bölgesel sistolik

değişiklikler, genellikle EKG'deki ST segment değişikliklerden ve göğüs ağrısının gelişiminden önce gelir.

İndüklenebilir duvar hareket anormalliklerinin varlığı, stresin zirve noktasında koroner arterlerde önemli bir kan akışı sınırlaması anlamına gelir ve genellikle %50 den fazla bir koroner arter darlığına işaret eder. Nispeten hafif koroner darlıklar durumunda iskeminin ortaya çıkabilmesi, maksimum stres performansına bağlıdır.

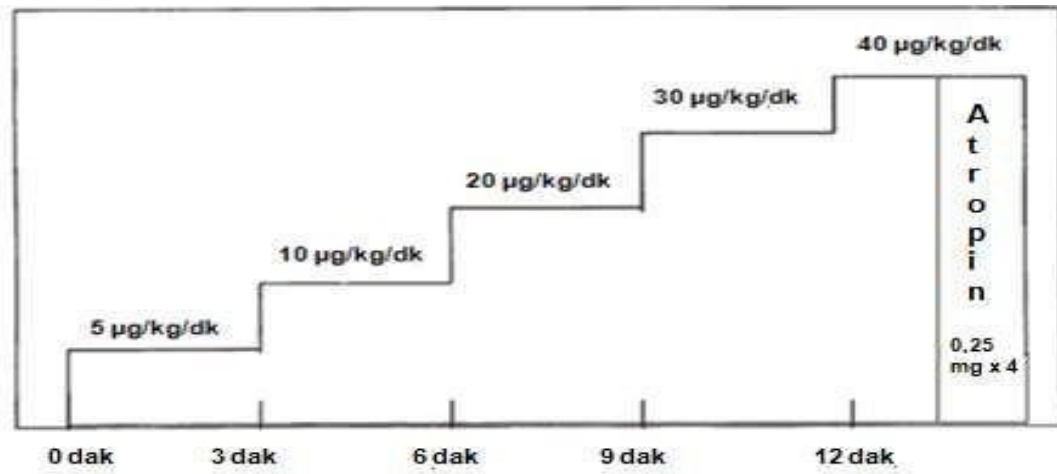
İndüklenebilir duvar hareketi anormallikleri genellikle stresten sonra hızla düzelir, ancak iskemi şiddetliyse ve stunning dene duruma neden olursa kalıcı olabilir.

2.5.2. Dobutamin Stres Ekokardiyografi

Dobutamin, sempatik beta reseptörler üzerinden etkili, sentetik dopamin analogudur.⁷⁴⁻⁷⁵ Dobutamin verildikten 2 dakika sonra etkisi başlamakta ve 10 dakika içinde maksimum etki görülmektedir. Dopaminin düşük dozları özellikle beta 1 ve beta 2 uyarımı yaparak pozitif inotrop ve kronotropik etkide bulunur. Ayrıca vazodilatasyon yapıcı etkisi ön plandadır. Bunun sonucunda kalp hızı, sol ventrikül atım volümü ve kardiyak output artar. Yüksek dozlardaki etkisi ise alfa 1 reseptörler üzerindendir ve bu etki ile arteriyal ve venöz vazokonstriksiyonda (damar duvarı düz kasılması) artış olur.⁷⁶ 5 pg/kg/dk dozunda sadece inotropik etkisi varken, 20 pg/kg/dk ve üzerindeki dozlarda kalp hızını artırarak kan basıncını ve miyokardın oksijen tüketimi artırmaktadır.^{70,76} Dobutamin ile yapılan stres testinin net etkisi egzersize (kalp hızıü kardiyak output, ve sistolik kan basıncında artış) çok benzemektedir. Bu durumda da miyokardın oksijen ihtiyacında artış olmaktadır. Eğer hastanın koroner arterlerinde ciddi darlık varsa, sol ventrikül kasının ihtiyacı olan kan sunumu karşılanamamakta ve sol ventrikül duvarında bölgesel iskemi olmaktadır, stres ile birlikte bu iskemi daha da aşikâr hale gelmektedir.⁷⁷⁻

2.5.2.1. Dobutamin Stres Ekokardiyografi Protokolü

Günümüzde Avrupa Ekokardiyografi Cemiyeti'nin (EAE) önerdiği standart protokolde dobutamin 5 $\mu\text{g/kg/dk}$ dozundan başlayarak her 3 dakikada bir dobutamin dozu artırılarak sırasıyla 10, 20, 30, 40 $\mu\text{g/kg/dk}$ dozlarına kadar çıkmak hedeflenmiştir. Yaşa göre maksimal kalp hızının % 85'i olan hedef kalp hızına ulaşamayan hastalarda atropin yapılabilir (0.25 miligramlık bölünmüş dozlarda her 1-3 dakika içinde bir kez maksimum 1 miligram).(Şekil 2.5.)^{81,82}



Şekil 2.5. Standard dobutamin stres ekokardiyografi protokolü

Ekokardiyografide sol ventrikülün herhangi bir duvar segmentinin akinetik olması her zaman skar dokusu veya kalıcı sol ventrikül fonksiyon bozukluğu anlamına gelmez. Hiberne veya sersemlemiş miyokard denen iki kavram mevcuttur. Bu durumlarda sol ventrikül miyokard dokusunun fonksiyonları damar açma işlemi sonrası normale geldiğinden, skar dokusundan ayırıcı tanısı yapılmalıdır. Miyokard canlılık değerlendirmesinde altın standart PET olmasına rağmen DSE tekrarlanabilir, kolay ulaşılabilir ve daha ucuz olması nedeniyle sıklıkla tercih edilmektedir. İstirahat anında yapılan ekokardiyografisinde miyokard dokusunun kalınlığının normal olması canlı miyokard, normalden ince ($< 6\text{mm}$) olması ve ekodan yoğun olması skar dokusunu düşündürmektedir.⁸³ DSE hem canlılık hemde iskemi araştırması şansı sunmaktadır. Eğer

bir hastada istirahat halinde sol ventrikül duvar hareketleri normal iken, dobutamin dozunun artması ile birlikte duvar hareket kusuru gelişıyorsa bu durum iskemiği gösteren bir bulgudur.

2.5.2.2. Dobutamin Stres Ekokardiyografi Yanıtları

Sol ventrikül canlılığı araştırılırken DSE de aşağıdaki dört yanıt görülür (Tablo 2. 3).^{82,83}

1. **Bifazik yanıt:** Bu tip yanıtta düşük doz dobutamin infüzyonu (5-10 pg/kg/dk) ile duvar hareketleri düzelir. Fakat dobutamin dozu arttıkça (20-30-40 pg/kg/dk) sol ventrikül duvar hareketleri kötüleşir. Bu yanıt miyokardın iskemik fakat canlı olduğunu ve damar açma işlemi sonrası sol ventrikül fonksiyonunu düzelme ihtimalinin yüksek olduğunu ifade eder.
2. **Giderek kötüleşen yanıt:** Bu grup hastalarda da damar açma işlem sonrası işlemde fayda gören gruplardandır. Burada duvar hareketlerine kötüleşme düşük dozlarda başlar ve doz arttıkça kötüleşme artar. Ciddi iskemi bulgusudur. Aynı zamanda canlılık göstergesidir.
3. **Giderek düzelen yanıt:** İstirahat halinde bozuk olan sol ventrikül duvar hareketlerinin düşük doz dobutamin ile düzelmesi ve doz arttıkça düzelmenin daha da belirgileşmesidir. Bu da canlı ve iskemik miyokardı işaret eder. uvar hareketlerinin düşük doz dobutamin ile iyileşmesi ve yüksek doza çıkıldığında aynı iyileşmeyi artırarak sürdürmesidir. Canlı ve iskemik olmayan miyokardı ifade eder. İşlem sonrası bu grupta düzelme beklentisi diğer iki gruba göre daha düşüktür.
4. **Sabit yanıt:** Dobutamin dozu arttıkça sol ventrikül duvar hareket kapasitesinde bir değişiklik olmaz. Bu durum skarı işaret eder.⁸³

Tablo 2.3. Dobutamin stres ekokardiyografinin yorumlanması

Başlangıç sol ventrikül fonksiyonu	Düşük doz dobutamin cevabı	Yüksek dobutamin cevabı	Sonuç	Geri dönüşüm olasılığı
Anormal	Düzelme var	Azalmış	Canlı ve iskemik (Hibernasyon)	Yüksek
Anormal	Bazalden kötü veya değişiklik yok	Bazalden veya düşük dozdan kötü	Canlı	Orta
Anormal	Düzelme var	Düzelme var.	Canlı ve iskemik değil (Stunning)	Düşük
Anormal	Düzelme yok.	Düzelme yok.	Skar	Düşük

2.5.2.3. Dobutamin İnfüzyonunu Durdurma Kriterleri

Test esnasında aşağıdaki durumlar gelişirse test sonlandırılır. Aksi halde hedef kalp hızına ulaşıncaya kadar dobutamin infüzyonunda devam edilir.⁷¹

1. Yeni oluşan veya kötüleşen duvar hareket kusuru.
2. Tolere edilemeyen yan etkiler gelişmesi
3. Baş ağrısı, anksiyete, dispne, bulantı, kusma, , tremor gibi semptomların çok belirgin olması
4. Kan basıncında sistolik kan basıncının 240 mmHg üzerine veya diyastolik kan basıncının 120 mmHg üzerine çıkması veya kan basıncının düşmesi ile birlikte hipotansiyon semptomlarının ortaya çıkması
5. EKG değişikliklerinin oluşması
6. Ciddi aritmilerin gelişmesi (ventriküler veya supraventriküler aritmiler)
7. Hedef doza ulaşılması (40 pg/kg/dk)

2.5.2.4. Hasta Monitörizasyonu

12 derivasyonlu EKG, bazal kalp hızı, kan basıncı, ve bazal ekokardiyografik görüntüler kayıt altına alınmalıdır. EKG kaydı test bitene kadar devam etmelidir. Kan basıncın ölçümü her aşamada ve hipotansiyon ya da hipertansiyon semptomları oluştuğunda ölçülmelidir. Hastalar dobutamin infüzyonunu takiben en az 5 dakika veya

vital bulgular normale gelinceye kadar (kalp hızı 100 atım /dakikanın altına düşünceye veya bazal değerdan en fazla 20 atım /dakika fazla oluncaya kadar, kan basıncı normale dönünceye, aritmiler sonlanıncaya kadar), semptomlar bitinceye, herhangi bir elektrokardiyografik veya ekokardiyografik iskemi bulgusu ortadan kalkana kadar monitörize edilmelidir.⁸⁴⁻⁸⁶

2.5.2.5. Dobutamin Stres Ekokardiyografi İçin Kontrendikasyonlar

1. Akut koroner sendromlar (kararsız angina, akut MI)
2. Dekompanse kalp yetersizliği
 1. Kontrolsüz taşikardiler
 2. Ciddi sol ana koroner arter darlığı bulunması
 3. Kan basıncı değerdanın 200 /110 mmHg üzerinde olması
 4. Hipertrofik obstrüktif kardiyomiyopati (Semptomatik)
5. Akut peri/miyokardit, endokardit
6. Semptomatik ileri aort darlığı
7. Glokom veya prostat hipertrofisi olan hastalarda atropin kullanımı.^{71,75}

3. MATERYAL VE METOT

3.1. Araştırma Türü

Araştırma ileri dönük (prospektif) bir çalışmadır. Çalışma 1999 yılında “Journal of American College of Cardiology” de yayınlanmış olup, veri seti bağımsız araştırmacıların kullanması için herhangi bir kısıtlama olmadan erişime açılmıştır. Biz de bu veri setini <https://hbiostat.org/data/repo/stressEcho.htm> adresinden indirdik.⁸⁷

3.2. Araştırmanın Yapıldığı Yer ve Zaman

Çalışma Mart 1991 ile Mart 1996 arasında, UCLA (University of California at Los Angeles) Tıp fakültesi Yetişkin Kardiyak Görüntüleme ve Hemodinami laboratuvarına sevk edilen hastalardan oluşmaktaydı.

3.3. Araştırma Popülasyonu

Mart 1991 ile Mart 1996 arasında, UCLA Tıp Fakültesi Yetişkin Kardiyak Görüntüleme ve Hemodinami laboratuvarına sevk edilen 1138 hasta alındı. Bu süre zarfında, hastada birden fazla DSE varsa, ilk ölçüm dikkate alındı. Ortotopik karaciğer nakli düşünülen 208 hasta, hastanede yatarken kalp dışı nedenlerden dolayı ölen 12 hasta, ortotopik kalp nakli yapılan 2 hasta ve DSE öncesi 6 ay içerisinde PTCA yapılan 13 hasta çalışmaya dahil edilmedi. Ayrıca, 12 aylık takibi olmayan 376 hasta da çalışmaya dahil edilmedi. Son olarak veri setinde pek çok değişkeni kayıp olan 14 hasta da çalışma dışı bırakıldı. Nihai çalışma popülasyonu 558 kişiden oluşmaktaydı.

3.4. Dobutamin Stres Ekokardiyografi

Dobutamin, üç dakika boyunca 5 mcg/kg/dk'dan başlayan ve her üç dakikada bir 5 mcg/kg/dk'lık artışlarla maksimum 40 mcg/kg/dk'ya kadar ilerleyen standart bir ölçülü uygulama cihazı kullanılarak intravenöz olarak uygulandı. Çalışmanın son iki yılında, ara 25- ve 35-mcg/kg/dk dozları ihmal edildi. Kalp hızı yanıtı yetersiz görüldüğünde testi yapan klinisyenin takdirine bağlı olarak atropin intravenöz olarak verildi. Test semptom

sınırlıydı ve hasta >2 mm ST depresyonu, şiddetli duvar hareketi anormallikleri, sık ventriküler veya supraventriküler ektopi, ventriküler taşikardi, supraventriküler taşikardi, sistolik kan basıncı >220 mm Hg, sistolik kan basıncının başlangıç sistolik kan basıncının 20 mm Hg altına düşmesi veya sistolik kan basıncının 80 mm Hg'nin altına düşmesi, bulantı, orta ila şiddetli göğüs ağrısı veya orta ila şiddetli dispne oluştuğunda durduruldu.

1 mm veya daha fazla düz veya aşağı eğimli ST depresyonu iskemi için pozitif kabul edildi. Bazal repolarizasyon anormalliklerinin varlığında, sol dal bloğu veya digoksin tedavisi alanlarda daha düşük derecelerde ST depresyonu veya ST değişiklikleri iskemi için şüpheli kabul edildi. Hasta sol lateral dekübit pozisyonunda iken istirahat ve dobutamin infüzyonunun her aşamasında ekokardiyografi yapıldı. Ek olarak, apikal 4 boşluk, apikal 2 boşluk, apikal uzun eksen ve parasternal kısa eksen görüntülerinin istirahat ve 5 mg/kg/dk, 10 mg/kg/dk ve pik dobutamin'de sayısallaştırılmış görüntüleri elde edildi. Miyokardın en iyi tanımına sahip görüntüler daha sonra aynı görüntüde sırayla görünen her bir görüntü için başlangıç, 5 mg/kg/dk, 10 mg/kg/dk ve pik dobutamin görüntüleri ile dörtlü ekran formatında sinelooplar olarak gösterildi. Miyokard, duvar hareketinin derecelendirilmesi için eşit parçalara bölündü. Üç apikal görünümdeki altı duvarın her biri (anterior, anterior interventriküler septum, lateral, inferior interventriküler septum, alt medial ve alt lateral) üç bölüme ayrıldı: bazal, orta ve apikal. Orta kısa eksen görünümü altı eşit parçaya bölündü ve apikal görünümde not edilen duvar hareketini doğrulamak için kullanıldı. Duvar hareketi her segment için hiperkinetik, normal, hafif hipokinetik, orta ila şiddetli hipokinetik veya akinetik/diskinetik olarak derecelendirildi. En az bir segmentteki duvar hareketi, dinlenme duvar hareketine göre en az bir derece bozulduysa, stres ekokardiyografik iskemi için pozitif kabul edildi. Akinezi ve diskinezi aynı derecede disfonksiyon olarak kabul edildi. Deneyimli bir

ekokardiyograf tarafından okunan nihai klinik raporda duvar hareketi için kullanılan yorumlar kullanılmıştır.

3.5. Araştırma Değişkenlerinin Tanımlanması

Tüm hastalarda, DSE sonrası takip eden 12 ay içinde aşağıdakilerden herhangi birinin meydana gelip gelmediğini belirlemek için tıbbi kayıt gözden geçirildi: “herhangi bir nedene bağlı ölüm, yeniden MI, takipte PTCA veya CABG ihtiyacı”. Analize yalnızca en az 12 aylık takibi olan veya ilk 12 ay içinde “herhangi bir nedene bağlı ölüm, yeniden MI, takipte PTCA veya CABG ihtiyacı” olarak tanımlanan bir kardiyak olayı olan hastalar dahil edildi.

Araştırma değişkenlerinin neler olduğu ve nasıl tanımlandığı aşağıdaki tabloda (Tablo 3.1) özetlenmiştir. Modellere dahil edilecek bağımsız değişkenler, daha önceki klinik tecrübe, biyolojik bilgi ve literatür taraması sonucu belirlendi.

Tablo 3.1. Araştırma değişkenlerinin tanımlanması

Bağımlı değişken (primer yanıt değişkeni):	
“herhangi bir nedene bağlı ölüm, yeniden MI, takipte PTCA veya CABG ihtiyacı”	YOK=0 VAR=1
Bağımsız değişkenler:	
Yaş,	Sürekli değişken(yıl)
Cinsiyet,	Erkek/Kadın (0/1)
Sistolik kan basıncı(SKB),	Sürekli değişken (mmHg)
DSE sonucu (+ veya -),	Pozitif/negatif (1/0)
Bazal LVEF,	Sürekli değişken (%)
Hipertansiyon,	YOK/VAR (0/1)
Diyabetes mellitus,	YOK/VAR (0/1)

3.6. İstatistiksel Analizler

Tüm istatistiksel analizler R-software v.3.5.1. (R statistical software, Institute for Statistics and Mathematics) ile yapılmıştır.

3.7. Araştırmada Kullanılan Analizler ve Modellerin Oluşturulması

Biz çalışmamızda, DSE sonrası 12-aylık takip süresi içerisinde gelişecek istenmeyen olayların (herhangi bir nedene bağlı ölüm, yeniden MI, takipte PTCA veya CABG ihtiyacı) tahmini için 3 farklı istatistiksel model kullandık:

1. Çok değişkenli lojistik regresyon (Multivariable logistic regression-MVLR)
2. Sınıflandırma ve regresyon ağaçları (Clasification and regression tree-CART)
3. Rasgele Orman Algoritması (Rastgele orman-RF)

3.8. Model Performanslarının Kıyaslanması

Daha sonra her bir model için internal validasyon yapıldı. Tercihen 200 bootstrap yeniden örnekleme ile her bir modelin R², brier skoru, AUC kaydedildi. Gözlemlenen olasılıklar ve model ile tahmin edilen olasılıklar için hem histogram eğrisi ile dağılımları verildi hem de “loess – locally weighted running line smoother” algoritması ile karşılıklı saçılım grafikleri “kalibrasyon” amaçlı çizdirildi.

4. BULGULAR

Tablo 4.1. Hastaların Temel Demografik, Klinik ve Laboratuvar Bulguları

Değişkenler		
Bağ	Cinsiyet (erkek, sayı ve %)	220(39.4)
ımsı	Yaş (yıl)	67.3+12.04
z	Diyabet (VAR, sayı ve %)	206(36,9)
deği	Hipertansiyon (VAR, sayı ve %)	393(70.4)
şke	KAH (VAR, sayı ve %)	210(37,6)
nler	Sigara (VAR, sayı ve %)	260(46.6)
	Göğüs ağrısı(VAR, sayı ve %)	172(30,8)
	Sistolik kan basıncı (mmHg)	146.9+36.5
	Bazal LVEF (%)	55+10.32
	Doputamin LVEF (%)	65.24+11.76
	İstirahat duvar hareket kusuru (VAR, sayı ve %)	257(46.1)
	Dobutamin stres testi pozitifliği(VAR, sayı ve %)	136(24.4)
Bağ	Kardiyak olay(VAR, sayı ve %)	90(16.1)
ımlı	Ölüm (VAR, sayı ve %)	24(4.3)
deği	Yeni MI (VAR, sayı ve %)	28(5)
şke	Yeni PTCA(VAR, sayı ve %)	27(4.8)
nler	Yeni CABG (VAR, sayı ve %)	33(5,9)

4.1. Bağımsız Değişkenler

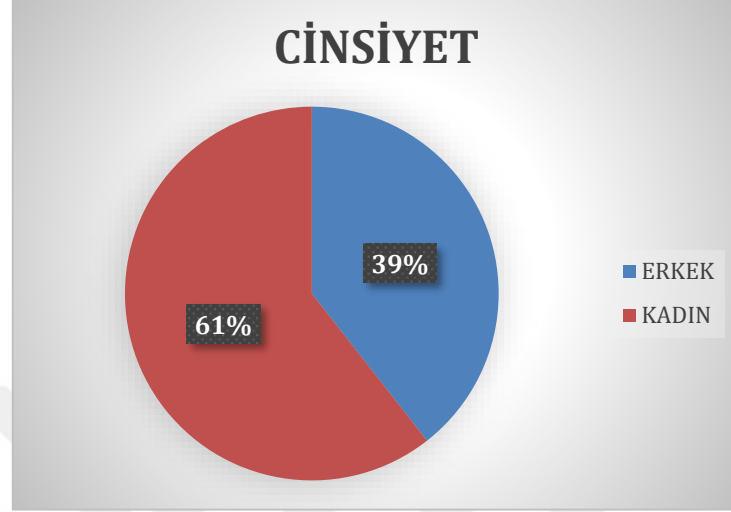
4.1.1. Kategorik Değişkenler

4.1.1.1. Cinsiyet

Tablo 4.2. Cinsiyet değişkeni için sayı ve yüzdeler

Cinsiyet	Sayı	Yüzde
KADIN	338	60.573
ERKEK	220	39.427
Total	558	100.000

Çalışmaya katılan hastalar cinsiyet dağılımlarına göre incelendiğinde 558 hastanın 220 tanesi erkek ve 338 tanesinin erkek olduğu görülmektedir. Hastaların cinsiyete göre dağılımı tablo ve şekilde gösterilmiştir.



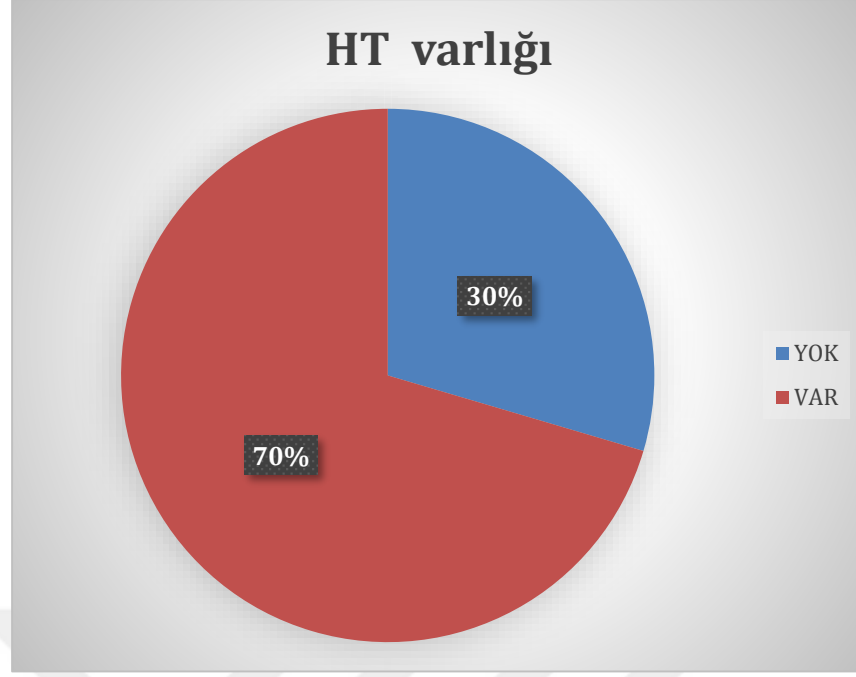
Şekil 4.1. Cinsiyet değişkeni için pasta grafiği

4.1.1.2. Hipertansiyon

Tablo 4.3. Hipertansiyon değişkeni için sayı ve yüzdeler

Hipertansiyon varlığı	Sayı	Yüzde
YOK	165	29.570
VAR	393	70.430
Total	558	100.000

Hipertansiyon varlığı kardiyovasküler hastalık riski için önemli bir risk faktörüdür. Toplumda oldukça yaygın görülmektedir. Bizim hasta popülasyonuzda hipertansiyon varlığı %70.4 olarak bulunmuştur. 558 hastanın 393'ünde hipertansiyon mevcuttur.



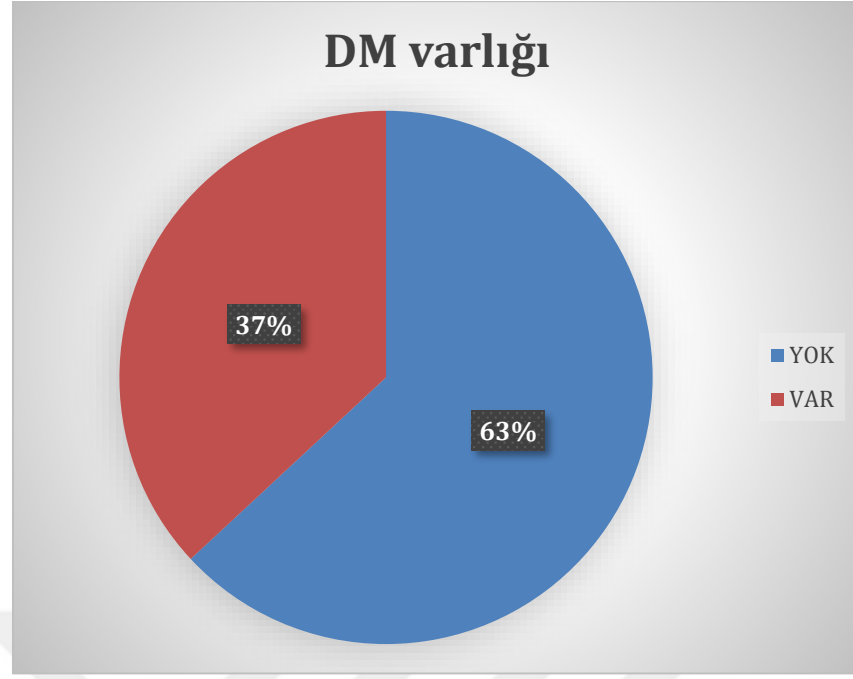
Şekil 4.2. Hipertansiyon değişkeni için pasta grafiği

4.1.1.3. Diyabetes Mellitus

Tablo 4.4. Diyabetes mellitus değişkeni için sayı ve yüzdeler

DM varlığı	Sayı	Yüzde
YOK	352	63.082
VAR	206	36.918
Total	558	100.000

Diyabetes mellitus kardiyak olay gelişimi için tanımlanmış en önemli risk faktörlerindendir. DM varlığında hem koroner arter hastalığı sık görülmektedir hem de olumsuz klinik sonuçları için önemli bir prediktördür. Bizim hasta grubumuzda DM oranı %36.9 olarak bulunmuştur. 558 hastanın 206'sında DM mevcuttur.



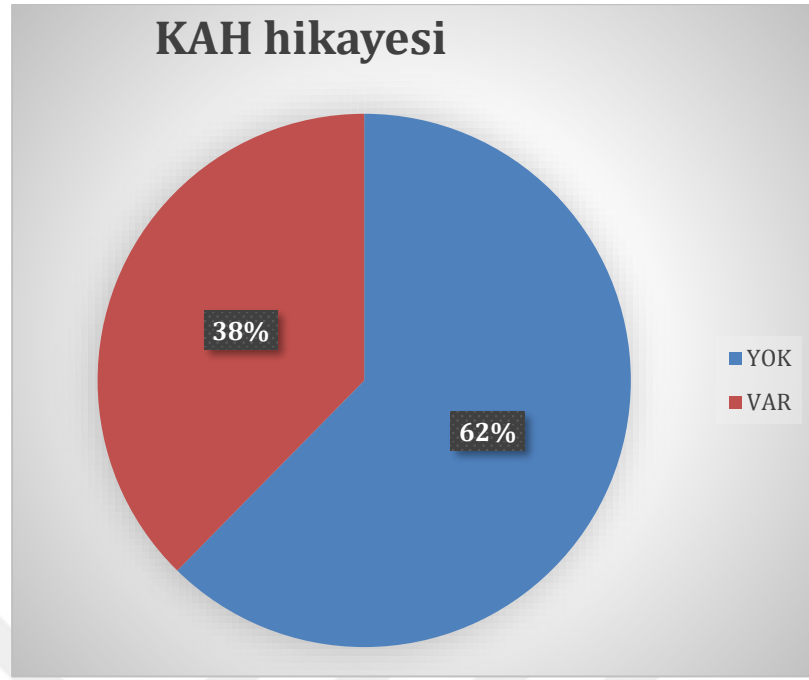
Şekil 4.3. Diyabetes mellitus değişkeni için pasta grafiği

4.1.1.4. Koroner Arter Hastalığı

Tablo 4.5. Koroner arter hastalığı değişkeni için sayı ve yüzdeler

KAH hikayesi	Sayı	Yüzde
YOK	348	62.4
VAR	210	37.6
Total	558	100.000

Hastanın hikayesinde herhangi bir koroner olay(MI, PTCA, CABG) yaşamış olması ilerleyen zamanlarda tekrar kardiyak yaşamasın predikte eden önemli bir değişkendir. Bizim hasta grubumuzun %37,6'sında KAH hikayesi mevcuttu.



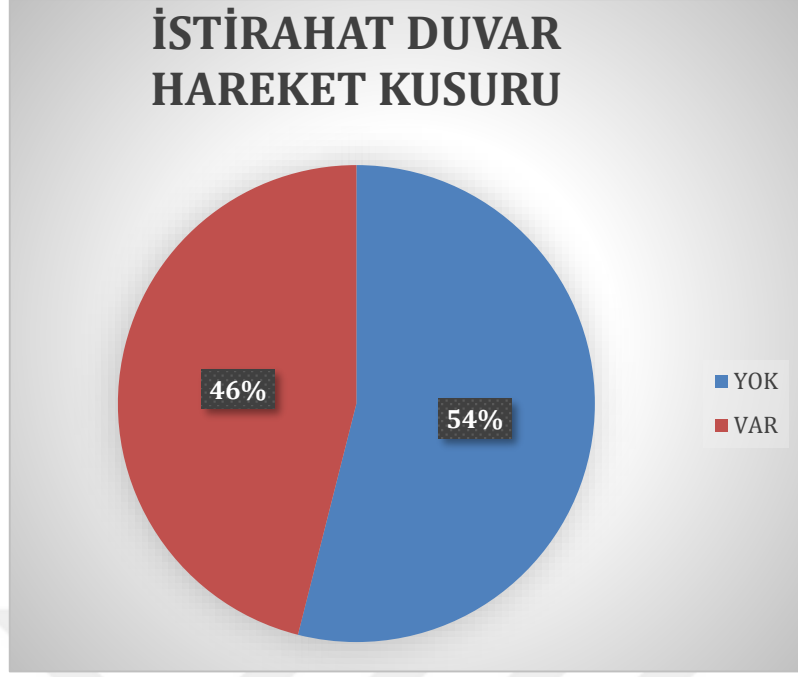
Şekil 4.4. Koroner arter hastalığı varlığı değişkeni için pasta grafiği

4.1.1.5. İstirahat Duvar Hareket Kusuru

Tablo 4.6. İstirahat duvar hareket kusuru değişkeni için sayı ve yüzdelere

İstirahat duvar hareket kusuru	Sayı	Yüzde
YOK	301	53.943
VAR	257	46.057
Total	558	100.000

Ekokardiyografide hastaların istirahat hareket kusuru varlığı kardiyak hastalık için önemli bir işarettir. Hastalarımızın istirahat duvar hareket kusuru varlığına göre değerlendirildiğinde %46 hastanın istirahatte duvar hareket kusurunu olduğu görülmektedir.



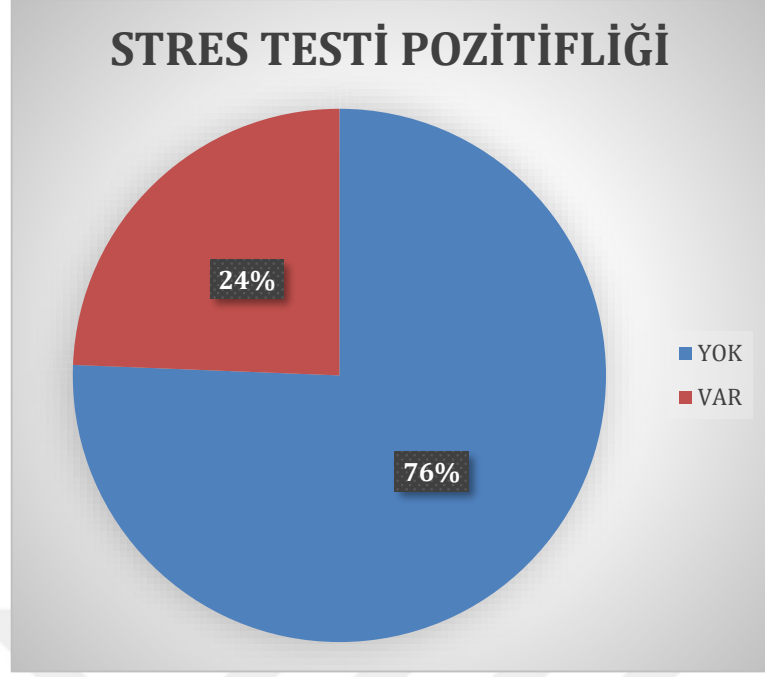
Şekil 4.5. İstirahat duvar hareket kusuru değişkeni için pasta grafiği

4.1.1.6. Stres Testi Pozitifliği

Tablo 4.7. Stres testi değişkeni için sayı ve yüzdeler

Stres testi pozitifliği	Sayı	Yüzde
YOK	422	75.627
VAR	136	24.373
Total	558	100.000

Çalışmanın primer başlangıç noktası olan dobutamin ile yapılan stres testi pozitifliği sonlanımlar üzerinde önemli etkisi olan bir parametredir. Bu parametre açısından kantitatif inceleme yapıldığında hastaların %24'ünde stres testinin pozitif olduğu tespit edilmiştir. Sayısal veriler tablo be şekilde gösterilmiştir.



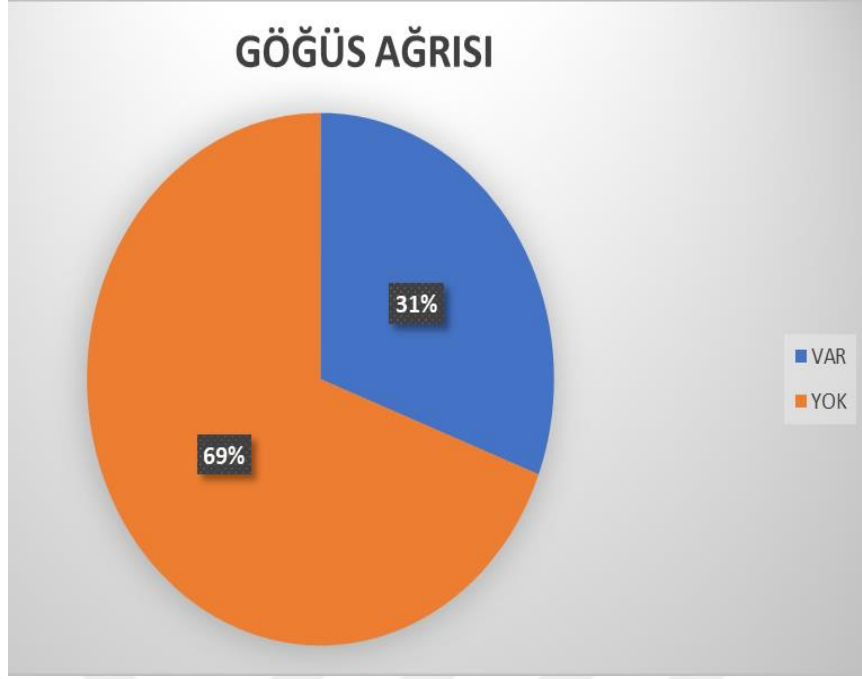
Şekil 4.6. Stres testi değişkeni için pasta grafiği

4.1.1.7. Göğüs Ağrısı Varlığı

Tablo 4.8. Göğüs ağrısı varlığı değişkeni için sayı ve yüzdeler

Göğüs ağrısı	Sayı	Yüzde
YOK	386	69.176
VAR	172	30.824
Total	558	100.000

Hastaların 386'sında göğüs ağrısı şikâyeti bulunmazken 172 hastanın başvuru şikâyeti göğüs ağrısı idi. Tablo ve şekilde sayı ve yüzdeler verilmiştir.



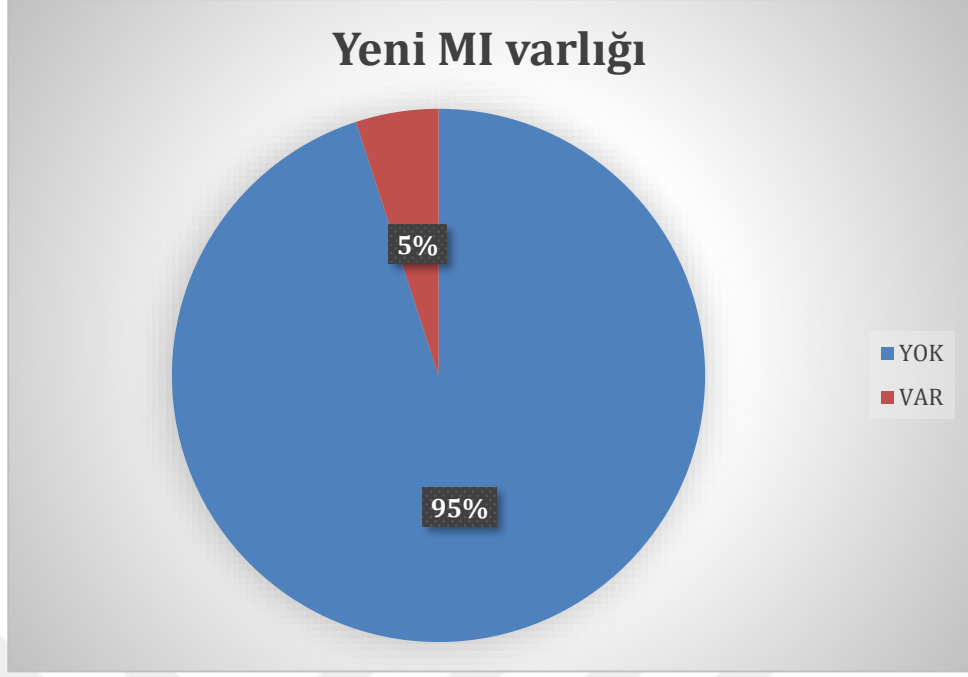
Şekil 4.7. Göğüs ağrısı varlığı değişkeni için pasta grafiği

4.1.1.8. Yeni MI Varlığı

Tablo 4.9. Yeni MI varlığı değişkeni için sayı ve yüzdeler

Yeni MI varlığı	Sayı	Yüzde
YOK	530	94.982
VAR	28	5.018
Total	558	100.000

Önemli bir klinik sonlanım olan yeni MI varlığı tüm hastalar içinde %5 hastada görülmüştür. Çalışmaya katılan 558 hastanın 28'inde MI sonlanımı gerçekleşmiştir.



Şekil 4.8. Yeni MI varlığı değişkeni için pasta grafiği

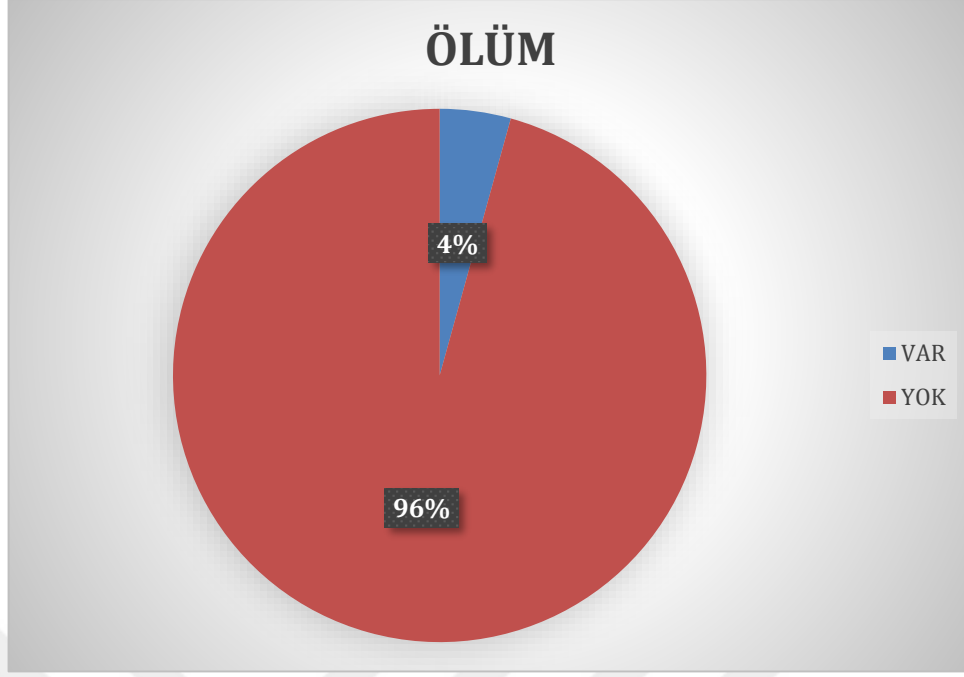
4.1.1.9. Ölüm

Tablo 4.10. Ölüm değişkeni için sayı ve yüzdeler

Ölüm	Sayı	Yüzde
YOK	534	95.699
VAR	24	4.301
Total	558	100.000

En önemli klinik sonlanım verisi olan ölüm tüm hastaların %4.3'inde görülmüştür.

Tabloda da görüldüğü üzere 558 hastanın 24 ü ölmüştür



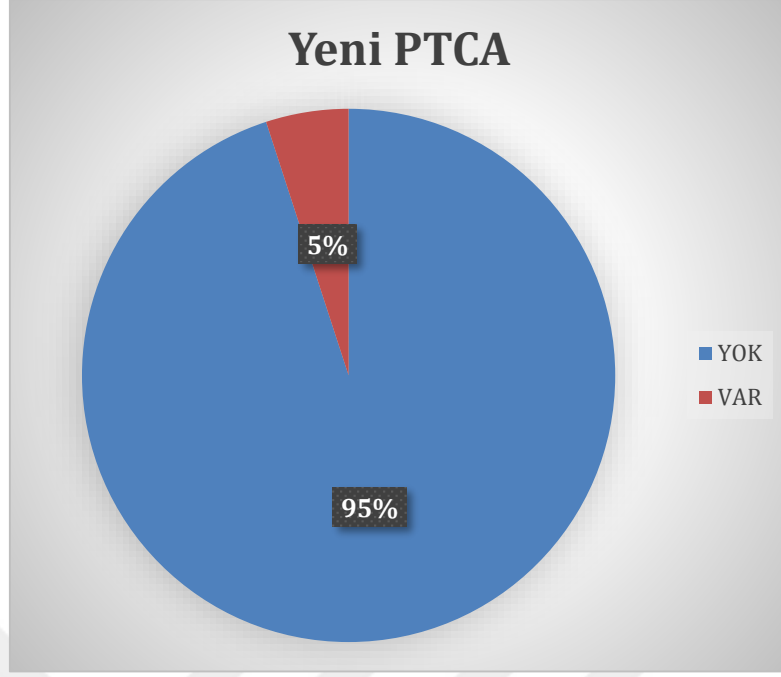
Şekil 4.9. Hipertansiyon değişkeni için pasta grafiği

4.1.1.10. Yeni PTCA İhtiyacı

Tablo 4.11. Yeni PTCA ihtiyacı değişkeni için sayı ve yüzdeler

Yeni PTCA	Sayı	Yüzde
YOK	531	95.161
VAR	27	4.839
Total	558	100.000

Hastaların yeni PTCA ihtiyacı olması önemli bir girişimsel sonlanımdır. 558 hastanın 27sinde yeni PTCA ihtiyacı olmuştur. Şekilde görüldüğü gibi bir sayı %4 hastaya tekabül etmektedir.



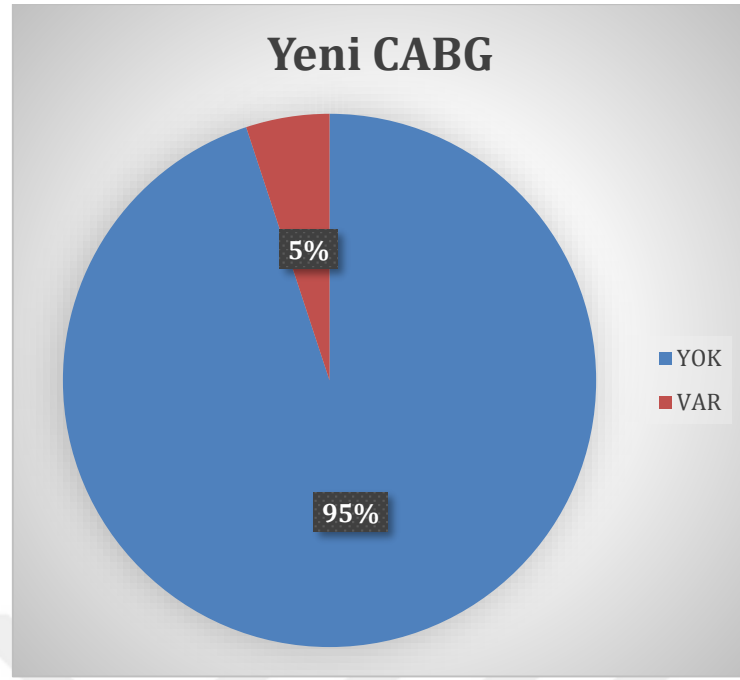
Şekil 4.10. Ölüm değişkeni için pasta grafiği

4.1.1.11. Yeni CABG İhtiyacı

Tablo 4.12. Yeni CABG değişkeni için sayı ve yüzdelere

Yeni CABG	Sayı	Yüzde
YOK	525	94.086
VAR	33	5.914
Total	558	100.000

Hastaların takipte CABG ihtiyacı olması anlamlı bir girişimsel sonlanım olarak değerlendirilmektedir. Bu önemli girişimsel sonlanım hastaların 33'ünde gözlenmiştir.



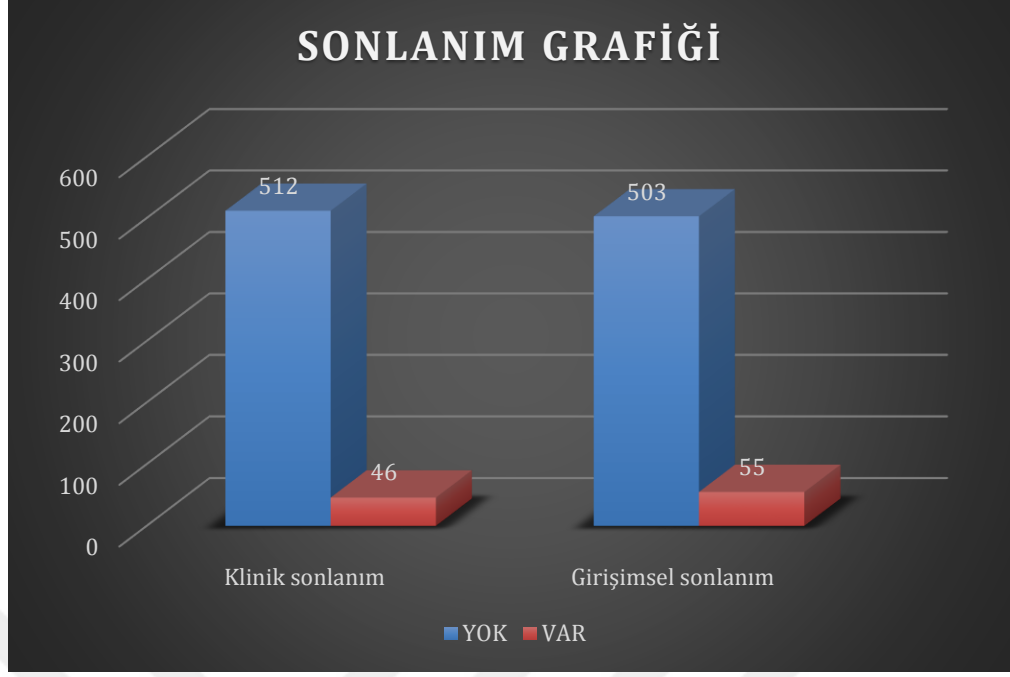
Şekil 4.11. Yeni CABG değişkeni için pasta grafiği

4.1.1.12. Klinik Sonlanım- Girişimsel Sonlanım

Tablo 4.13. Klinik sonlanım ve girişimsel sonlanım değişkenleri için sayı ve yüzdelere

	Klinik sonlanım varlığı		Girişimsel sonlanım varlığı	
	Sayı	Yüzde	Sayı	Yüzde
YOK	512	91.756	503	90.143
VAR	46	8.244	55	9.857
Total	558	100.000	558	100.000

Hastalarda ölüm ve MI gelişmesi klinik sonlanım, yeni PTCA veya yeni CABG gelişmesi ise girişimsel sonlanım olarak belirlenmiştir. Hastaların %8.2 sinde klinik sonlanım gelişirken, %9.8'inde girişimsel sonlanım gelişmiştir.



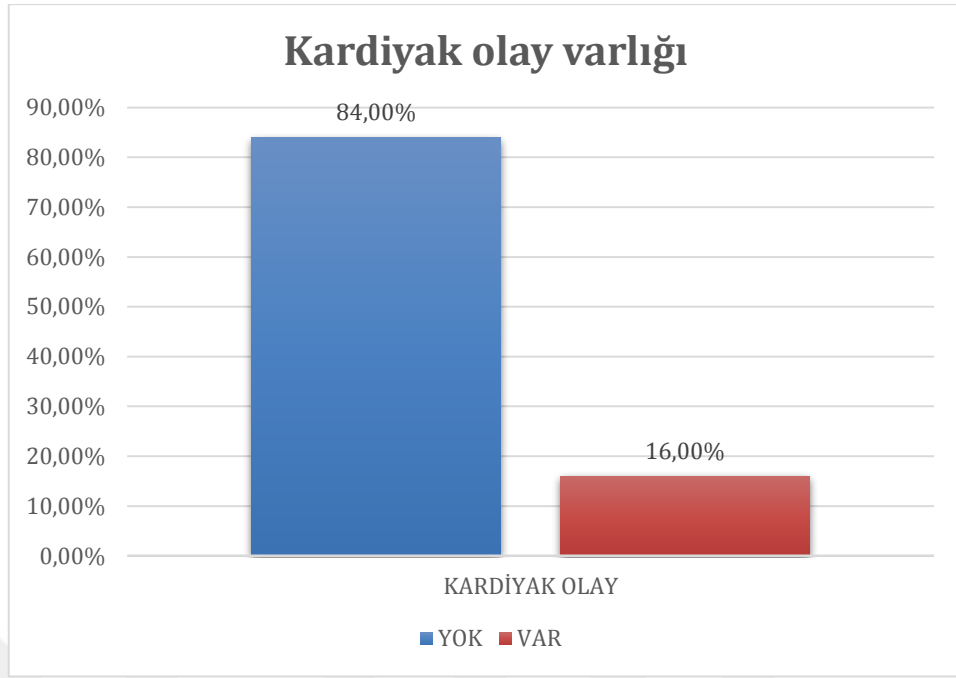
Şekil 4.12. Klinik sonlanım ve girişimsel sonlanım değişkenleri sütun grafiği

4.1.1.13. Kardiyak Olay

Tablo 4.14. Kardiyak olay gelişmesi değişkeni için sayı ve yüzdeler

Kardiyak olay varlığı	Sayı	Yüzde
YOK	468	83.871
VAR	90	16.129
Total	558	100.000

Herhangi bir nedene bağlı ölüm, takipte MI gelişmesi, takipte PTCA veya CABG ihtiyacı olması kardiyak olay olarak tarif edilmiştir. Tüm hastalar değerlendirildiğinde kardiyak olay gelişen hasta sayısı 90'dır. Bu sayı toplam hasta sayısının %16.12'sini oluşturmaktadır.



Şekil 4.13. Kardiyak olay varlığı için sütun grafiği

Kardiyak olay ve gelişmeyen hastalarda bağımsız değişkenlere ait sayısal veriler tabloda verilmiştir.

Tablo 4.15. Kardiyak olan gelişen ve gelişmeyen gruplarda değişkenlerin karşılaştırılması

Değişkenler	KARDİYAK OLAY (+)	KARDİYAK OLAY(-)
Cinsiyet (erkek, sayı ve %)	43(47.7)	177(37.8)
Yaş (yıl)	69.14±9.60	66.99±12.44
Diyabet (VAR, sayı ve %)	45(50)	161(34.40)
Hipertansiyon (VAR, sayı ve %)	74(82.22)	319(68.16)
Göğüs ağrısı (VAR, sayı ve %)	138(29.48)	330(70.51)
Sistolik kan basıncı (mmHg)	141.66±34.76	147.92±36.80
Bazal LVEF (%)	50.26±12.96	56.63±9.40
Doputamin LVEF (%)	58.03±15.30	66.62±10.41
İstirahat duvar hareket kusuru (VAR, sayı ve %)	15(16.66)	242(51.70)
Dobutamin stres testi pozitifliği (VAR, sayı ve %)	47(52.22)	89(19.01)

4.1.2. Sayısal Değişkenler

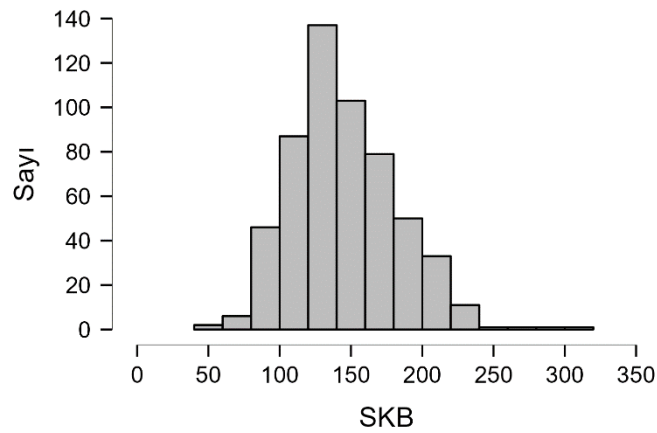
Çalışmada beş farklı sayısal bağımsız değişken incelenmiştir. Bunlar yaş, sistolik kan basıncı (SKB), dobutamin dozu, bazal LVEF ve dobutamin LVEF sürekli sayısal değişkenlerdir. Tabloda sayısal değişkenlere ait ortalama, standart sapma, minimum ve maksimum değerleri verilmiştir.

Tablo 4.16. Sayısal değişkenlere ait veriler

	Sistolik KB	Dobutamin Dozu	Yaş	Bazal EF	Dobutamin EF
Ortalama	146.916	30.242	67.344	55.604	65.242
Standart sapma	36.530	9.535	12.049	10.322	11.762
Minimum	40.000	5.000	26.000	20.000	23.000
Maksimum	309.000	40.000	93.000	83.000	94.000

4.1.2.1. Sistolik Kan Basıncı

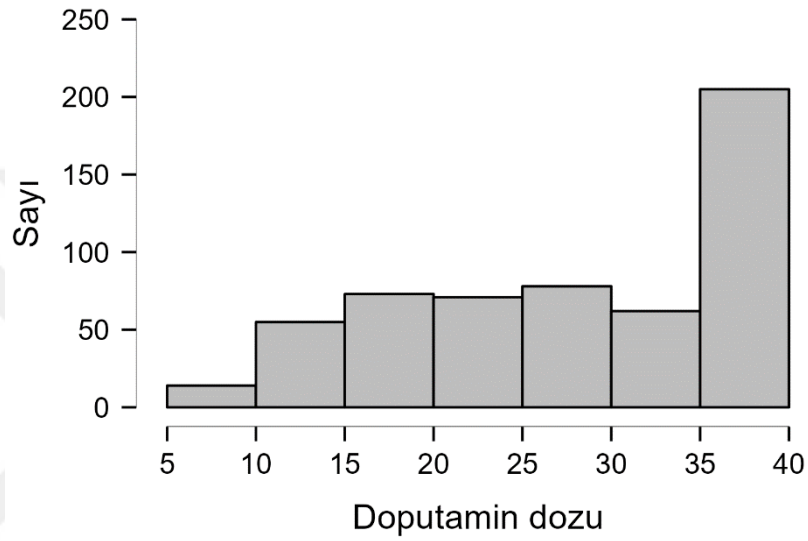
Sistolik kan basıncı önemli bir kötü prognostik göstergedir. SKB arttıkça prognoz kötüleşmektedir. Hastalarımızı SKB değerleri ortalama 146,9 mmHg olarak hesaplanmıştır. Maksimum değeri 309 ve minimum değeri 40 mmHg dir. SKB nin dağılım grafiği Şekil de gösterilmiştir. Grafikten de anlaşılacağı üzere SKB değişkeni normal dağılım göstermektedir.



Şekil 4.14. Sistolik kan basıncı için verilen dağılım grafiği

4.1.2.2. Dobutamin Dozu

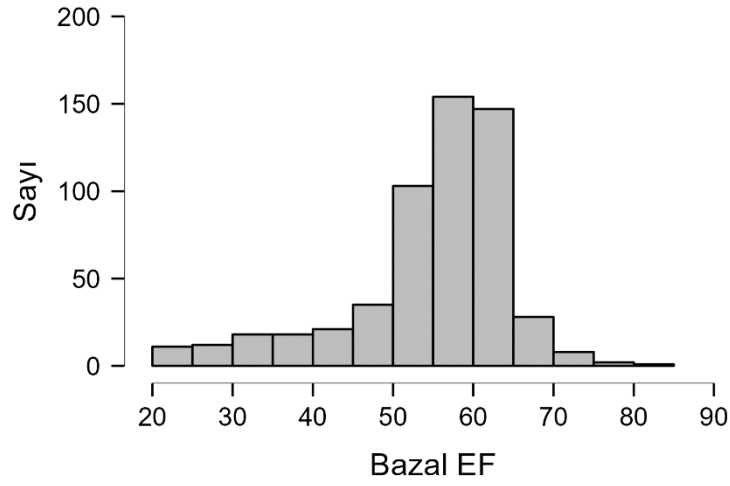
Dobutamin yavaş yavaş artırılan dozlarda verilerek sol ventrikül ejeksiyon fraksiyonunu artışı beklenmektedir. Eğer doz artımına bağlı ejeksiyon fraksiyonu artışı yoksa sonlanımlar açısından olumsuz bir durum mevcut demektir. Hastalarımıza ortalama 30,2 mikrogram dozunda doputamin infüzyonu verilmiştir. Doz dağılım grafiği şekilde gösterilmiştir. Grafikten görüleceği üzere dobutamin dozunda dağılım normal değildir.



Şekil 4.15. Dobutamin dozu için verilen dağılım grafiği

4.1.2.3. Bazal Ejeksiyon Fraksiyonu

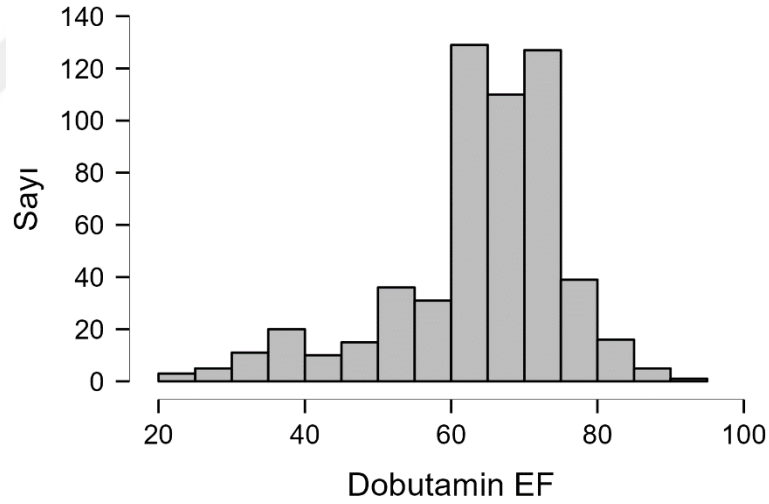
Bazal EF için ortalama 55,6 olarak bulunmuştur. Minimum değeri 20 ve maksimum değeri 83 olarak hesaplanmıştır. Bazal EF için dağılım grafiği şekilde gösterilmiştir.



Şekil 4.16. Bazal EF için verilen dağılım grafiği

4.1.2.4. Dobutamin Ejeksiyon Fraksiyonu

Dobutamin EF için ortalama 65.24, standart sapması 11.762 dir. Maksimum değer 23 ve minimum değer 94 olarak hesaplanmıştır. Dağılım grafiği şekilde gösterilmiştir.



Şekil 4.17. Dobutamin EF için verilen dağılım grafiği

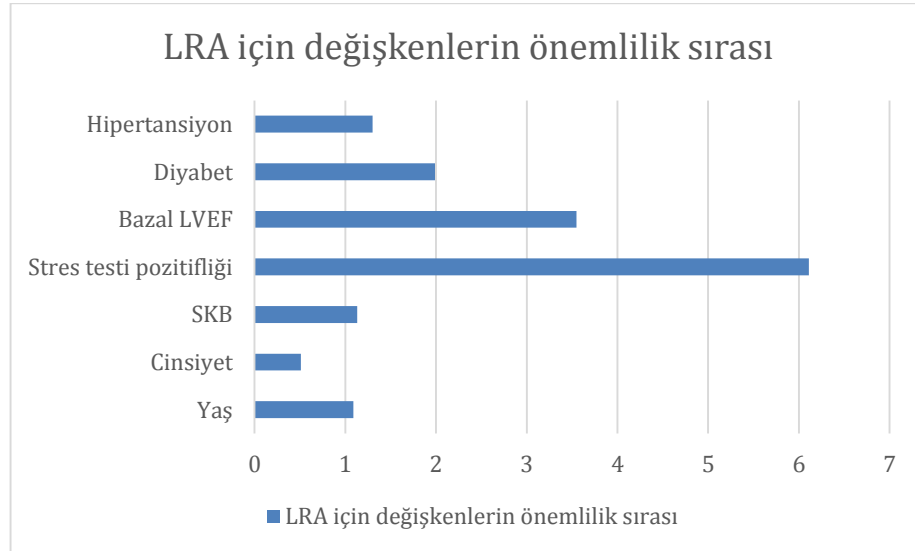
4.2. Değişkenlerin Önemlilik Sırası

Değişkenlerin önemlilik sırası tablo 4.17.'de özetlenmiştir.⁸⁸

Tablo 4.17. Değişkenlerin önemlilik sırası

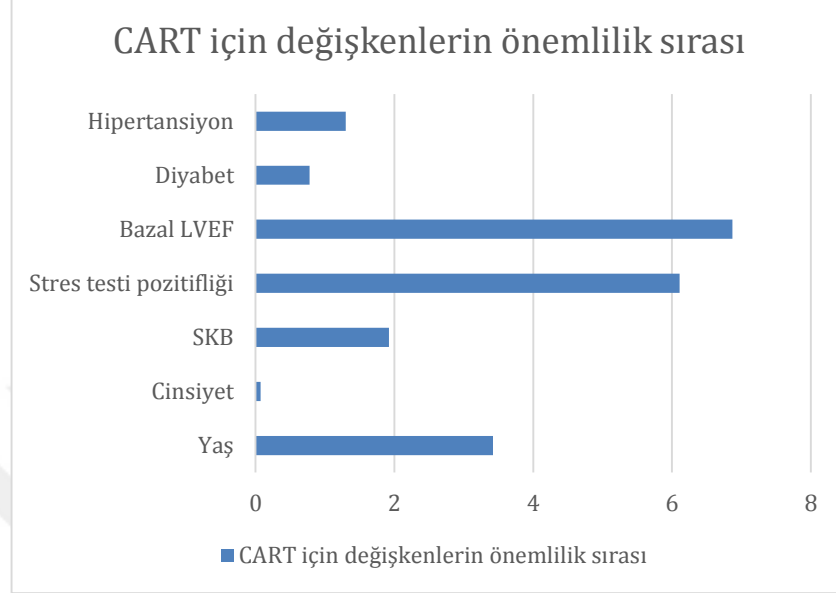
	CART	Lojistik regresyon	Rastgele orman
Yaş	3.42	1.09	0.92
Cinsiyet	0.07	0.51	4.24
SKB	1.92	1.13	-3.23
Stres testi pozitifliği	6.11	5.32	11.6
Bazal LVEF	6.87	3.55	17.8
Hipertansiyon	1.30	2.34	3.26
Diyabet	0.78	1.99	3.39

Değişkenlerin önemlilik sırası kullanılan analiz yöntemine göre değişiklik göstermektedir. Şekilde de görüldüğü gibi lojistik regresyon analizinde en önemli parametre stres testi pozitifliği iken onu bazal LVEF ve diyabet takip etmektedir. Önemlilik sırası en geride olan değişkenin ise cinsiyet olduğu gözükmemektedir. (şekil)



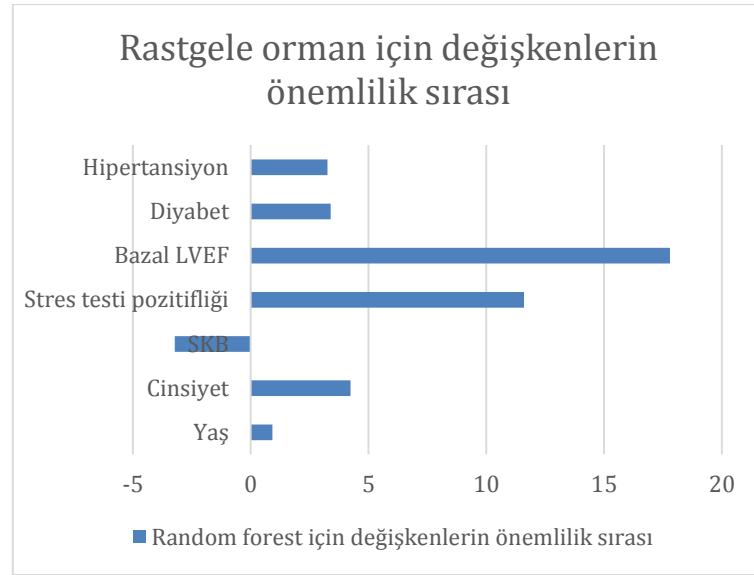
Şekil 4.18. Logistik regresyon analizi için değişkenleri önemlilik sırası grafiği

CART analizinde önemlilik sırasında ilk üç değişken bazal LVEF, stres testi pozitifliği ve yaş değişkenleridir. Cinsiyet ise LRA da olduğu gibi en önemsiz değişkendir.(Şekil)



Şekil 4.19. CART için değişkenleri önemlilik sırası grafiği

Rastgele orman ile model oluşturulurken bazal LVEF en önemli değişken olarak karşımıza çıkmaktadır. Onu sırasıyla stres testi pozitifliği ve cinsiyet izlemektedir. Diğer analizlerde farklı olarak cinsiyet değişkeni rastgele orman için önemli bir değişken olarak karşımıza çıkmaktadır. (Şekil 4.20.)



Şekil 4.20. Rastgele orman için değişkenleri önemlilik sırası grafiği

4.3. Model Performansların Karşılaştırılması

Model performanslarını karşılaştırılması için kullanılan birçok ölçüt bulunmaktadır. Biz tezimizde performans karşılaştırması için R², Brier skoru, ölçekli brier, AUC (C istatistiği), kalibrasyon intercept, kalibrasyon slope ve ECI (estimated kalibrasyon indeksi) kullandık.

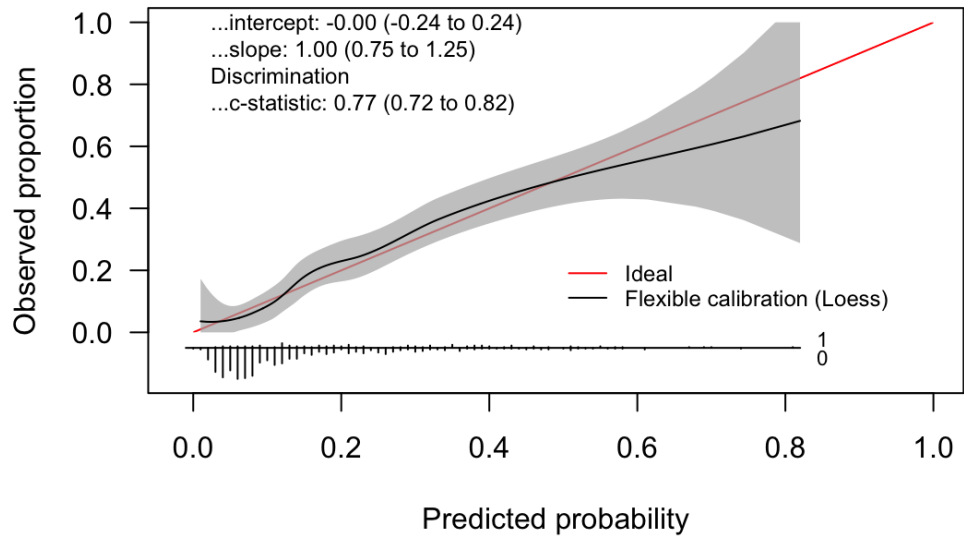
Tabloda performans ölçütleri özetlenmiştir.

R² ne kadar yüksekse model o kadar iyidir. Tablodan da görüldüğü gibi R² değeri CART için 0.283'dir ve en yüksektir. Dolayısıyla performans ölçütü en iyi olan CART analizidir. Rastgele orman için R² değeri 0.118'dir ve performansı en düşük analiz rastgele orman'dır. Brier skoru R²'nin aksine ne kadar düşükse performans o kadar iyidir. Tabloda görüldüğü gibi brier skoru en düşük olan CART analizidir ve en iyi performans göstergesi ona aittir. AUC(C istatistiği) ayırt edicilik göstergesidir. Bir model olay gelişenlerle olay gelişmeyenleri ne kadar iyi ayırt ediyorun göstergesi AUC'dir. Tabloda görüldüğü gibi AUC değeri en yüksek olan CART analizidir. AUC CART analizi 0,785 dir ve logistik regresyon ve rastgele orman için sırasıyla 0,772, 0,704'dür.

ECI değeri ne kadar düşükse performans o kadar iyidir. Tabloda görüldüğü gibi CART performansı ön plana çıkmaktadır. Kalibrasyon eğrisinde intercept 0 ve slope 1 ise model performansı o kadar iyidir. Bu değerler CART ve logistik regresyon için geçerli iken, rastgele orman bu şartları sağlamamaktadır.

Tablo 4.18. CART, LRA ve RF için model performans ölçütlerinin karşılaştırılması

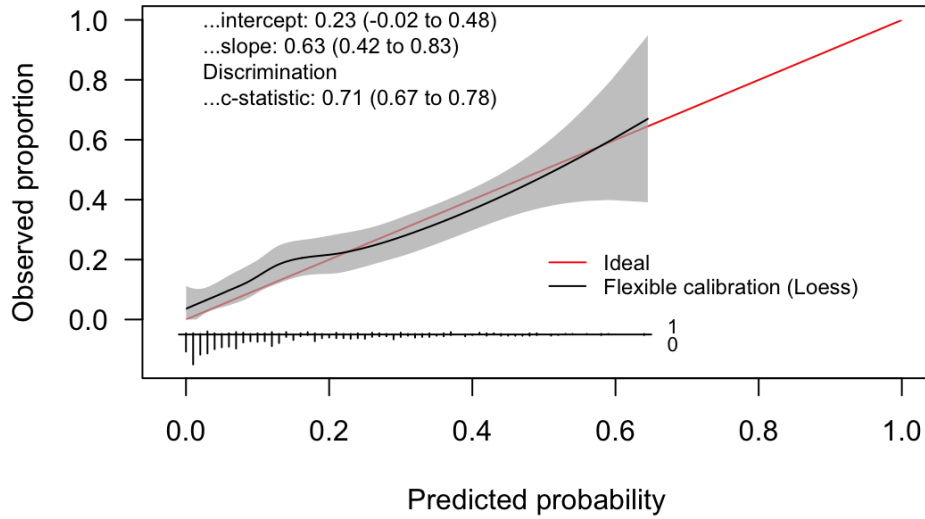
	CART	Lojistik regresyon	Rastgele orman
R²	0.283	0.208	0.118
Brier	0.105	0.116	0.132
Brier scaled	0.227	0.138	0.273
AUC (C istatistiği)	0.785	0.772	0.704
Kalibrasyon intercept	0.00	0.00	0.260
Kalibrasyon slope	1.00	1.00	0.587
ECI	0.00	0.048	0.159



Şekil 4.21. Logistik regresyon analizi için kalibrasyon eğrisi

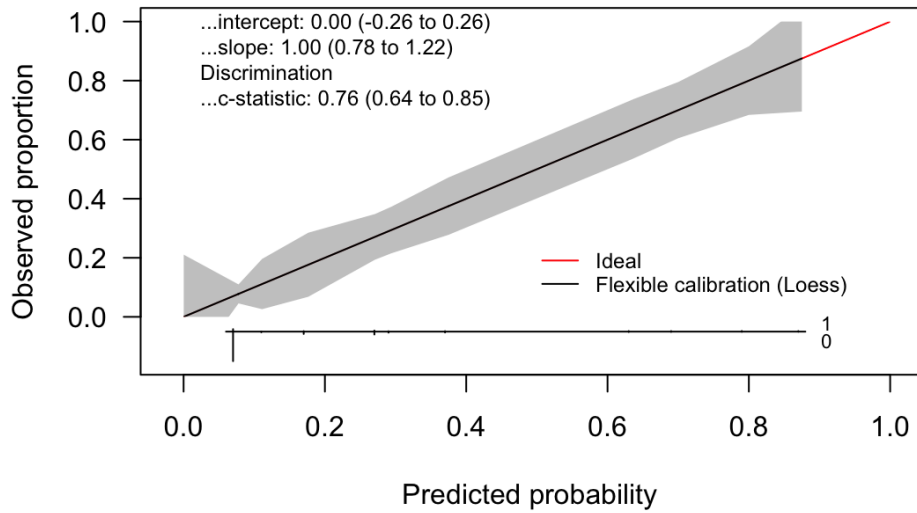
Kalibrasyon eğrisinde kırmızı çizgi ile gösterilen diyagonal çizgi tam uyumu gösterir. Siyah çizgi logistik regresyon için Loes yöntemi ile çizdirilen kalibrasyon eğrisini göstermektedir. Eğridende anlaşılacağı gibi tahmin edilen risk % 50 üzerinde iken modelin underestimate yaptığı görülmektedir. Intercept 0 ve slope 1 olması ile ideal bir model özelliği göstermektedir. (Şekil 4.20)

Şekil 4.21’de rastgele orman için Loes yöntemi ile çizilen kalibrasyon eğrisi görülmektedir. Kırmızı çizgi tam uyumu gösteren ideal diyagonal çizgidir. Siyah çizgi rastgele orman için çizilen kalibrason çizgisidir. Tahmin edilen risk düşük olduğu zaman modelin overestimate yaptığı grafikten gözükmemektedir. Intercept 0 ve slope 1 olmadığı için ideal bir model olmadığı anlaşılmaktadır.



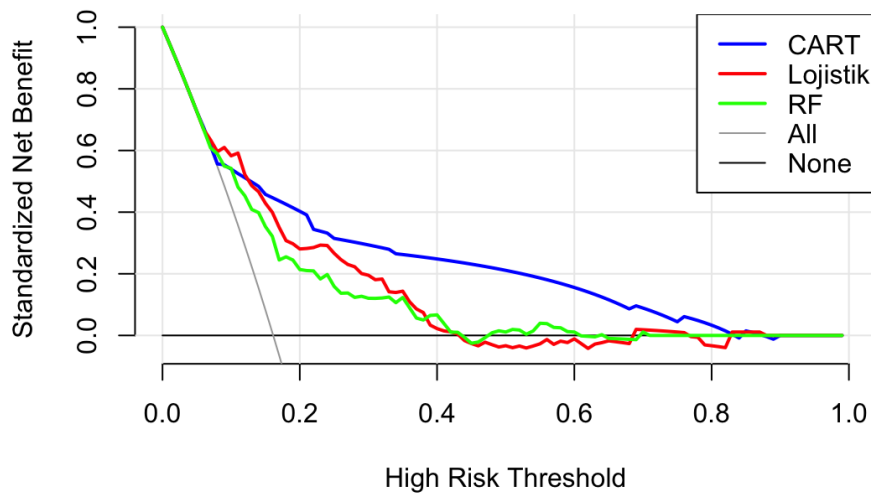
Şekil 4.22. Rastgele orman algoritması için kalibrasyon eğrisi

Şekilde CART analizine ait kalibrasyon eğrisi görülmektedir. Görüldüğü üzere ideal diyagonal eğri ile tam uyumlu bir kalibrasyon eğrisi izlenmektedir. Bu da CARTın karşılaştırılan diğer yöntemlere göre daha ideal bir tahmin modeli oluşturduğunu göstermektedir. Zaten diğer performans ölçütleri de bunu desteklemektedir.

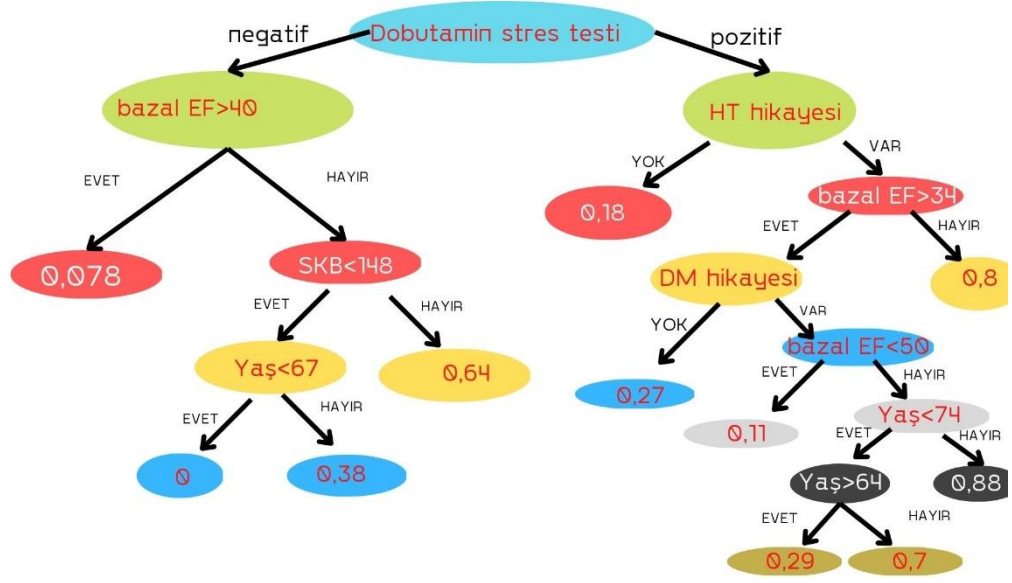


Şekil 4.23. CART için kalibrasyon eğrisi

Bir modelin klinik karar verme gücünü göstermek için en sık kullanılan ölçüt desicion cuve analizidir(DCA). Daha önceden belirlenmiş bir olasılık aralığında model veya modellerin default(varsayılan) yöntemlere göre net klinik faydasını gösterir. DCA eğrisinde x ekseninde olasılık eşiği, y ekseninde net klinik fayda gözükmektedir. Şekilde mavi ile gösterilen CART için verilen DCA klinik karar verme konusunda en iyi modelin CART ile oluşturulduğunu göstermektedir. (Şekil 4.23)



Şekil 4.24. CART, lojistik regresyon analizi ve Rastgele Orman için DCA eğrisi



Şekil 4.25. Klinik sonuç için CART ağacı

5. TARTIŞMA

Bu tez çalışmasında CART, logistik regresyon analizi ve rastgele orman modellerinin dobutamin stres ekokardiyografi sonuçlarını yorumlama güçleri ve model performansları karşılaştırılmıştır. Çalışmamızda; 8 prediktör her üç regresyon modeline dahil edilerek modellerin güçleri ile ölüm, MI, bypass ve PTCA öngördürücülükleri karşılaştırılmıştır. Her üç model de iyi düzeyde performans göstermiştir. CART en iyi model performansı sağlarken rastgele orman en düşük düzeyde model performansı sağlamıştır.

Klinik çalışmalarda araştırmacıların yaptığı iş bağımsız değişkenlerle bağımlı değişken arasında ilişki olup olmadığı ve bu ilişkinin klinik yansıması değerlendirilmesidir. Bundan dolayıdır ki çalışma dizayn edilirken istatistiksel metodun seçilmesi ve sonrasında çalışmanın yapılması temel kuraldır. Özellikle randomize çalışmalarda çalışma protokolü ve istatistiksel metod ilk tanımlanan parametrelerdir. Bağımlı değişkenimiz ejeksiyon fraksiyonu gibi sürekli bir değişken ise doğrusal regresyon modeli karşımıza çıkarken, diyabet ya da hipertansiyon gibi kategorik bir değişken olduğunda ise normal dağılım ilkesi bozulacağı için logistik regresyon modelleri karşımıza çıkmaktadır. Lojistik regresyonda, normallik varsayımının bozulduğu için lineer regresyon analizine alternatif olarak kullanılabilir. Her iki regresyon modelinde ki temel farklılık ise lineer regresyon analizinde yapılan işi bağımlı değişkenin alacağı değer tahmin edilmeye çalışılır. Lojistik regresyon analizinde ise yapılan iş bağımlı değişkenin alma ihtimali olan değerlerden birinin gerçekleşme ihtimalini bulmaktır.

Son yıllarda kardiyoloji alanındaki gelişmelere paralel olarak kardiyoloji dergilerinin sayısı ve makale sayısı artış göstermiştir. Kardiyovasküler hastalıkların uzun dönem etkileri yada verilen tedavinin etkileri etraflıca incelenmiş ve incelenmeye devam edilmektedir. Tüm bunların sonucunda hem prognostik faktör arayışı hemde advers olay

tahmininde parametre arayışı istatistiksel metodların kullanımını kardiyoloji alanında yaygınlaştırmıştır.

Bu bağlamda en fazla kullanılan analiz metodu lojistik regresyon analizidir. Lojistik regresyon analizinin kullanım kolaylığı ve yorumlanmasının kolay olması nedeniyle kardiyoloji alandında son zamanlarda çok sık tercih edilen bir istatistiksel metod olarak karşımıza çıkmaktadır.

İster randomize isterse non-randomize çalışmalarda kardiyovasküler olay tahmini kritiktir ve yapılan büyük çaplı çalışmaların çoğusunda regresyon analizi kullanılmıştır.⁸⁹⁻

⁹² Elenie ve arkadaşları atriyal fibrilasyon hastalarında oral antikoagulan yoğunluğunun stroke ve mortaliteye etkilerini araştırmış ve INR >2 olan hastalarda hem stroke hem de stroke nedeni ölümde azalma olduğunu lojistik regresyon modeli ile ortaya koymuştur.⁹³ Jonh ve arkadaşları ise yüksek volümlü anjiyoplasti merkezlerinde mortalite oranlarının daha düşük olduğunu regresyon modeliyle ortaya koymuştur.⁹⁴ Son olarak gilles ve arkadaşları ise perkütan girişimlerde enoksaparin ile klasik heparin etkilerini regresyon analizi ile karşılaştırmış ve IV enoksaparin tedavisinin alternatif olabileceğini öne sürmüştür.⁹⁵

Bilimsel çalışmaların sonucu ortaya çıkan sonuçların analizinde son zamanlarda CART kullanım alanı olarak sıklıkla kendine yer bulan bir istatistiksel metod olarak karşımıza çıkmaktadır. Çalışılan veri setinin hiçbir varsayım ihtiyacı olmaması nedeniyle, sınıflama ve regresyon ağaçları (CART) klasik regresyon modellerine karşı avantajlı ve değerli bir alternatif olarak ortaya çıkmıştır. CART kategorik ve sürekli değişkenleri bir arada kullanarak sınıflama ve regresyon sorunlarının çözümünde karar ağaçlarını kullanan parametrik olmayan istatistiksel bir yöntemdir. Veri olarak kullanılan bağımlı değişken kategorik ise metod sınıflama ağaçları (Classification Tree), eğer bağımlı değişkenimiz sürekli bir değişken ise regresyon ağaçları (Regression Tree) olarak

adlandırılmaktadır. Bu tanımdan da anlaşılacağı üzere CART, çoklu regresyon analizini hem de bağımlı değişkenimizin kategorik olduğu durumlarda kullandığımız lojistik regresyon analizini içinde barındırmaktadır.²⁵

İstatistiksel araştırmalarda kullanılan CART algoritması, her basamakta ilgili kümeyi bir üst kümeden daha homojen olan iki alt kümeye bölerek ikili karar ağaçları ortaya çıkaran bir dizayna sahiptir. Burada yapılmak istenen, hedef değişken hakkında mümkün olabilecek en homojen veri alt gruplarını oluşturmaktır. CART, yalnızca bağımlı değişken ile bağımsız değişken arasındaki ilişkinin biçimini irdelemekle kalmayıp, ek olarak bağımsız değişkenlerin kendi aralarında olan etkileşimlerini de araştırmayı amaçlayan bir yöntemdir. CART yönteminin, bağımsız değişkenlerin bağımlı değişkenle olan ilişkisini ortaya koymada ve model yapısındaki ilişki yapısını çözümlemede çok önemli ayrıcalıkları bulunmaktadır. CART'ın temelini oluşturan varsayım, birbirlerine benzerlik gösteren değişkenlerin aynı alt ağaç düğümünde birikmesine dayalı olup, bu yöntemle oluşturulan, alt dalları bağımlı değişkenimizle başlayan kök düğüme bağlayarak son bulmaktadır.⁴⁴

Peacock ve arkadaşları kalp yetersizliği hastalarında artmış kardiyak troponin düzeylerinin mortaliteyle olan ilişkisini CART ile göstermiştir.⁹⁶ Loise ve arkadaşları litik tedavi sonrası girişim gerekliliğinin prediktörlerini lojistik regresyon ve CART modeli ile beraber araştırmış ve CART modelinin sonuçla ilişkili ana faktörleri daha iyi tanımladığını, regresyonun ise birden fazla etkenin sonuçlarını aynı anda değerlendirmeye imkan verdiğini öne sürmüştür.⁹⁷ Janine ve arkadaşları ise dobutamin stres ekokardiyografi ile advers olay tahmini için CART analizi uygulamışlar ve pozitif SE'nin yanı sıra stres EF, istirahat duvar hareketleri, stres ECG yanıtında dikkate alınması gerektiği belirtilmiştir.⁸⁷ Çalışmamızda CART modeli bu çalışmaya benzer olarak daha iyi model performansı gösterdi.. Lojistik regresyon eş zamanlı tüm

parametreleri değerlendirirken CART parametreleri kümelerine ayırarak öngördürücü değişkenleri sırayla değerlendirir ve bir sonraki değişkenin seçimi önceki değişkenlerin sonuçlarına bağlıdır. Sonuç olarak CART bölünen dallanma kategorilerinin bir akış şemasını üretir ve hastaları farklı kardiyak olay olasılıklarına sahip olma durumuna göre alt gruplara ayırır. Bu ayırımı ise CART'ın gücünün temel kaynağıdır ve geleneksel regresyona olan üstünlüğü çalışmamızda da gösterilmiş olması literatür ile uyumludur. Ayrıca çalışmamızda CART modelindeki parametrelerin önemlilik sırası sırasıyla yaş, DSE ve bazal LVEF parametreleridir. CART modelinde DSE(6.11) ve Bazal LVEF(6.87) önemlilik sırasında ilk iki sırayı almıştır. Literatür incelendiğinde her iki parametrenin de aterosklerotik kardiyovasküler hastalıkların takip ve prognozunda önemli olduğu ve prognostik faktör olarak kullanılabileceği geniş çaplı çalışmalarda ortaya konmuştur. Çalışmamızda ki CART modelinin daha iyi performans ölçütü sunmasında ki temel etkenlerden biride bu durum olabilir.

Logistik regresyon modelleri kardiyovasküler sistem üzerine yapılan advers olay yada prediktör araştırma çalışmalarında en yaygın kullanılan istatistiksel metoddur. Diyabet yada hipertansiyon gibi nonparametrik değişkenleri modele dahil edilebilmesi onun yaygın olarak kullanılmasında ki temel etkidir. Zira kardiyovasküler olaylarda etkisi gösterilen yaş ve sistolik kan basıncı gibi parametrik değişkenler olduğu gibi diyabet, hipertansiyon, obezite gibi nonparametrik değişkenler de etkilidir. Bu parametrelerin hangi oranda kardiyovasküler hastalıklara etkili olduğu ya da riski ne kadar artırdığı araştırmacının ilk hedefidir ve bu durumda lojistik regresyon ilk akla gelen yöntemdir.³¹ Goligher ve arkadaşları COVID 19 hastalarında heparin tedavisinin kritik hastalarda mortaliteye etki etmediğini lojistik regresyon ile göstermişlerdir.⁹⁸ Mancia ve arkadaşları COVID 19 hastalarında ACEI ya da ARB kullanımının daha fazla olduğunu ancak bunun ilave risk getirmediğini lojistik regresyon modeli ile ortaya koymuştur.⁹⁹

Bizim çalışmamızda regresyon modelimizdeki parametrelerin önemlilik sırası sırasıyla HT(2.3), DSE(5.3) ve bazal LVEF(3.5) parametrelerinden oluşmaktaydı. Her üç parametrede advers olaylarla etkileri iyi gösterilmiştir. Modelimizin etkinliği literatür ile uyumludur.

Breiman ve arkadaşları tarafından yeni bir yöntem olarak tanımlanan Rastgele orman yöntemi hem sınıflandırma yöntemine hem de neden-sonuç ilişkisine dayanan ve regresyon amaçlı da kullanılabilen denetimli bir makine öğrenmesi yöntemlerinden birisidir. Rastgele orman birden fazla karar ağacını kullanıp daha uyumlu modeller oluşturarak daha isabetli tahminlerde bulunmaya yardımcı olan bir regresyon yöntemidir. Karar ağaçlarını kullanmasından dolayı belli bir aralıkta istenen tahminler için aynı sonuçları üretir yani kesikli değişkenlerdir.⁶³

Rastgele orman yönteminin avantajları olarak; hiperparametre kestirimi olmadan da daha uygun sonuçlar üretmesi, hem regresyon hem de sınıflandırma sorunları için kullanılabilir olmasından dolayı son zamanlarda sık tercih edilen makine öğrenmesi yöntemleri olmuşlardır.

Karar ağaçlarının en büyük problemlerinden biri olan overfitting problemi rastgele orman çözülmüş gibi gözükmetedir. Rastgele orman modeli overfitting problemini çözmek için rastgele onlarca belki yüzlerce farklı altkümeler seçer. Bu şekilde yüzlerce karar ağacı oluşturulur ve her bir karar ağacı bir tahminde bulunur. Rastgele orman modelinde farklı veri setleri üzerinde model eğitildiği için varyans, diğer bir tabirle karar ağaçlarının en büyük problemlerinden olan overfitting azalır.

Rastgele orman algoritması sınıflandırma sorunlarında genel geçer bir gerçek olan çoğunluğun haklılığı prensibine göre sınıflandırma yapar. Yani en çok tercih edilen, yani en çok oyu alan sınıf Rastgele orman'ın tahminini ortaya koymaktadır. Regresyon

sorunlarında ise Rastgele orman'ın tahmini ormanı oluşturan bulunan bütün ağaçların ortalaması alınarak bulunur ve bir anlamda ismini de bu tanımdan almıştır.

Venepally ve arkadaşları Nonbacterial thrombotic endocarditis prediktörlerini araştırdıkları çalışmada rastgele orman modelini kullanmışlar ve hastane içi mortalitenin yüksek olduğunu belirtmişlerdir¹⁰⁰. Ayrıca bu hasta grubunda ileri yaş, pulmoner emboli, sistemik emboli, mitral yetersizliği, antifosfolipid antikor varlığı da hastane içi mortaliteyi artıran durum olarak tespit edilmiştir. Walther ve arkadaşları kronik böbrek yetmezliği hastalarında vücut kitle indeksinin kalp yetmezliği benzeri semptom gelişimindeki öneminin rastgele orman ile ortaya koymuştur.¹⁰¹ Rastgele orman modelindeki parametrelerin önem sırası sırası ile revaskülarizasyon hikayesi (8.4), DSE (17.8) ve bazal LVEF (11.6) parametrelerinden oluşmaktadır. Her üç parametrede kötü kardiyak sonlanımla yakında ilişkilidir. Rastgele orman modelinin de iyi performans ölçütü sunmasında ki etken de bu durumdan kaynaklanma ihtimali vardır.

6. SONUÇ VE ÖNERİLER

- Regresyon modelleri kardiyovasküler hastalıkların tanısında tedavisinde ve advers olay gelişimi tahmininde güvenle kullanılabilir.
- Dobutamin stres ekokardiyografi CART modelinde ikinci, logistik regresyon modelinde birinci, rastgele orman modelinde ikinci önemli parametre olarak bulunmuştur.
- Bazal LVEF CART modelinde birinci, logistik regresyon modelinde ikinci, random forest modelinde birinci önemli parametre olarak bulunmuştur.
- Hipertansiyon logistik regresyon modelinde üçüncü önemli parametre olarak bulunmuştur.
- Revaskülarizasyon hikayesi rastgele orman modelinde üçüncü önemli parametre olarak bulunmuştur.
- Yaş CART modelinde ikinci önemli parametre olarak bulunmuştur.
- Diyabet rastgele orman modelinde dördüncü önemli parametre olarak bulunmuştur.
- HT tüm çalışma grubunda yüksek oranda tespit edilmiştir.
- Tüm çalışma grubunun %16'sında kardiyak olay gelişmiştir.
- Çalışmamızda en iyi model preformansı CART modeli ile elde edilmiştir.
- Çalışmamızda en kötü model preformansı rastgele orman modeli ile elde edilmiştir.
- DSE ve LVEF her üç regresyon modelinde ilk sıralarda kendisine yer edinmiştir.
- DSE ve LVEF'nin klinik öngördürücülükleri yüksektir.
- Kardiyovasküler olay tahmininde geleneksel regresyon modellerine alternatif olarak CART modeli güvenle kullanılabilir.

Çalışmamızın sınırlılıklarının ön önemlisi olarak belirteceğimiz husus ise çalışma grubumuzun hasta sayısı modelimizin klinik kullanımı için yeterli değildir. Buna karşın sonuçlarımız literatürde yer alan çalışmalarla benzerlik göstermektedir. Mortalite ve

advers olay görölme sayısı yine literatürdeki çalışma verileri ile benzerdir. Sonuç olarak hem çalışmamızın model etki gücü yüksektir, hem de her üç regresyon modelinin performansını karşılaştırması bakımından değerlidir ve kardiyovasküler sonlanımları öngörme açısından ileriki çalışmalara yol gösterici olabilir. Hatta hasta sayısı artırılarak aynı temel üzerine yapılan modellemeler klinik uygulamalarda yol gösterici olabilir.



KAYNAKLAR

1. Kalaivani M, Amudhan S, Upadhyay AD, Kamal VK(2017) , “Biostatistics”. Hemanshu Prabhakar(Ed.)Essentials of Neuroanesthesia (ss.979-995) London, Academic Press-Elsevier
2. Gacar BK, Kocakoç İD Regresyon Analizleri mi Karar Ağaçları mı?, *Manisa Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, 2020,18; 251-260.
3. Saitoğlu Y. Sınıflama Ve Regresyon Ağaçları. Doktora Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, 2015.
4. Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd Ed. New York: Springer. 2009, 311.
5. Linoff, Gordon S., and Michael J. A. Berry. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. 3rd ed. Indianapolis: Wiley Publishing, Inc., Chapter, 2011. 7.
6. Tuffery, Stephane. Data Mining and Statistics for Decision Making. Chichester, West Sussex, UK: John Wiley & Sons. 2011, 329.
7. Wu, Xindong; Kumar, Vipin; Ross Quinlan, J.; Ghosh, Joydeep; Yang, Qiang; Motoda, Hiroshi; McLachlan, Geoffrey J.; Ng, Angus; Liu, Bing; Yu, Philip S.; Zhou, Zhi-Hua "Top 10 algorithms in data mining". *Knowledge and Information Systems*. 2008. 14: 1-37.
8. Rokach, Lior; Maimon, O. Data mining with decision trees: theory and applications, 2nd Edition. World Scientific Pub Co Inc. 2014.
9. Berry, M. J., Linoff, G. S., “Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management 2nd ed.”, Wiley, USA, 2004.

10. Kıran Z. “Lojistik regresyon ve C&RT Analizi Yöntemleriyle Sosyal Güvenlik Kurumu İlaç Provizyon Sistemi Üzerinde Bir Uygulama”, Yüksek Lisans Tezi, Ankara, Gazi Üniversitesi, 2010.
11. Aydoğan, F. “ E-Ticarete Veri Madenciliği Yaklaşımlarıyla Müşteriye Hizmet Sunan Akıllı Modüllerin Tasarımı ve Gerçekleştirimi”, Ankara, Hacettepe Üniversitesi, 2003.
12. Altıntaş, Y. “Veri Madenciliğinin Tıpta Kullanımı Ve Bir Uygulama : Hemodiyaliz Hastaları İçin Risk Seviyelerine Göre Risk Faktörlerinin Etkileşimlerinin İncelemesi”, Yüksek Lisans Tezi, Ankara, Gazi Üniversitesi, 2010.
13. Köse O. Sınıflama Ve Regresyon Ağaçları Tekniği İle Kalp Hastalıklarına Etki Eden Bazı Faktörlerin Belirlenmesi, Yüksek Lisans Tezi, İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü, 2018.
14. Breiman, Leo; Friedman, J. H.; Olshen, R. A.; Stone, C. J. Classification and regression trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software. ISBN 978-0-412-04841-8, 1984.
15. Friedman, J. H. Stochastic gradient boosting [https://astro.temple.edu/~msobel/courses/files/StochasticBoosting\(gradient\).pdf](https://astro.temple.edu/~msobel/courses/files/StochasticBoosting(gradient).pdf). Stanford University, 1999.
16. Hastie, T., Tibshirani, R., Friedman, J. H. The elements of statistical learning Data mining, inference, and prediction. New York: Springer Verlag, 2001.
17. Breiman, L. "Bagging Predictors" Machine Learning, 1996. 24: 123,-140.
18. Rodriguez, J. J.; Kuncheva, L. I.; Alonso, C. J. "Rotation forest: A new classifier ensemble method". IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006, 28: 1619—1630.

19. Hothorn, T.; Hornik, K.; Zeileis, A. "Unbiased Recursive Partitioning: A Conditional Inference Framework". Journal of Computational and Graphical Statistics, 2006, 15: 651-674.
20. Strobl, C.; Malley, J.; Tutz, G. "An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forests" Psychological Methods, 2009. 14: 323-348.
21. Gareth, James; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert An Introduction to Statistical Learning New York: Springer. 2015, 315-323.
22. Bramer, M. Principles of Data Mining. London, Springer, 2007, 41-50
23. Ben-Gal I. and Trister C. "Parallel Construction of Decision Trees with Consistently Non Increasing Expected Number of Tests" (<http://www.eng.tau.ac.il/~bengal/Trist.pdf>) Applied Stochastic Models in Business and Industry, 2015, 31 64-78.
24. Deng, H.; Runger, G.; Tuv, E. Bias of importance measures for multi-valued attributes and solutions ([https:// www.researchgate.net/publication/221079908](https://www.researchgate.net/publication/221079908)). Proceedings of the 21st International Conference on Artificial Neural Networks, 2011, 293-300.
25. Güner, Z. B. "CART and Logistic Regression Analysis in Data Mining: An Application on Pharmacy Provision System Data" . Sosyal Güvenlik Uzmanları Derneği, *Sosyal Güvenlik Dergisi*, 2014,6.
26. Dondurmacı, G. Veri Madenciliğinde Regresyon Ağaçları ile Sınıflandırma ve Bir Uygulama Doktora Tezi, İstanbul, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü, 2011.
27. Tatlıdil, H. *Uygulamalı Çok Değişkenli İstatistiksel Analiz*, Ankara: 1. Basım, Cem Web Ofset, 2002.

28. Khemphila, A., Boonjing, V. Comparing performances of logistic regression, decision trees, and neural networks for classifying heart disease patients. International Conference on Computer Information Systems and Industrial Management Applications, CISIM 2010. 10.1109/CISIM.2010.5643666.
29. Hosmer, D. W., & Lemeshow, S. “Applied Logistic Regression”, John Wiley & Sons, New York, 1989, 5-50.
30. Eygü, H. Çözümlü Güncel Örneklerle Olasılık ve İstatistik. Nobel Yayıncılık. Birinci Baskı, Ankara, 2021.
31. Harrell FE. Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001:13-44
32. Zurada J., Lonial S., “Comparison Of The Performance Of Several Data Mining Methods For Bad Debt Recovery In The Healthcare Industry”, *The Journal of Applied Business Research*, 2005, 21: 37-53.
33. Hand, J., “Data mining: statistics and more ?”, *The American Statistician*, 1998, 52: 112-118.
34. Altıntaş, Y. “Veri Madenciliğinin Tıpta Kullanımı Ve Bir Uygulama : Hemodiyaliz Hastaları İçin Risk Seviyelerine Göre Risk Faktörlerinin Etkileşimlerinin İncelemesi”, Yüksek Lisans Tezi, Gazi Üniversitesi, 2010.
35. Maimon R. “Data Mining & Knowledge Discovery Handbook” ,Springer, 2005, 334.
36. Han J., Kamber M., “Data Mining Concepts and Techniques 2nd Ed.”, Editor : Jim Grey, The Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann, 2006, 2398-403
37. Tatlıdil H. “Uygulamalı Çok Değişkenli İstatistiksel Analiz”, Engin Yayınları, Ankara, 1992, 11, 122, 252-260

38. Kuyucu Y. Lojistik Regresyon Analizi , Yapay Sinir Ağları Ve Sınıflandırma Ve Regresyon Ağaçları Yöntemlerinin Karşılaştırılması Ve Tıp Alanında Bir uygulama. Yüksek Lisans Tezi, Tokat, Osmangazi Üniversitesi Sağlık Bilimler Enstitüsü, 2012.
39. Ocakoğlu G., “Lojistik Regresyon Analizi ve yapay Sinir Ağları Yöntemlerinin Sınıflama Özelliklerini Karşılaştırılması ve Bir Uygulama”, Yüksek Lisans Tezi, Uludağ Üniversitesi, Bursa, 2006.
40. Kurt, İ. Türe, M. Kurum, A. T. Comparing Performances of Logistic Regression, Classification and Regression Tree, and Neural Networks for Predicting Coronary Artery Disease. . Expert Systems With Applications, 2008, 34, 366-374.
41. Kocakoç, İ., & Keser, İ. Exploring Decision Rules for Election Results by Classification Trees. In Economies of the Balkan and Eastern European Countries, Kne Social Sciences, Pages 107-115. DOI 10.18502/Kss.V4i1.5982economies Of The Balkan and Eastern European Countries (EBEEC 2019), Conference Paper. 2019.
42. Lewis, R. J. “An Introduction to Classification and Regression Tree (CART) Analysis” Ucla Medical Center Torrance, California Presented at the 2000 Annual Meeting Of The Society For Academic Emergency Medicine In San Francisco, California, 2000.
43. Deconinck, E, Hancock, T., Coomans, D., & Massart. “Classification of Drugs in Absorption Classes Using the Classification and Regression Trees (CART) Methodology”, Journal of Pharmaceutical and Biomedical Analysis, 2005, 39: 91-103.
44. Breiman, L. Friedman, J., Olshen, R. ve Stone, CJ, Sınıflandırma ve Regresyon Ağaçları. 1987.

45. Dolgun, M. Ö. Veri Madenciliği Sınıflama Yöntemlerinin Başarılarının; Bağımlı Değişken Prevelansı, Örneklem Büyüklüğü ve Bağımsız Değişkenler Arası İlişki Yapısına Göre Karşılaştırılması, Doktora Tezi, Ankara: Hacettepe Üniversitesi, 2014.
46. Gökay Emel, G. ve Taşkın, Ç. Pazarlama Stratejilerinin Oluşturulmasında Bir Karar Destek Aracı: Birliktelik Kuralları Madenciliği. Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 2005. 7, 30-59.
47. Larose, DT. Verilerdeki Bilgiyi Keşfetmek: Veri Madenciliğine Giriş. New Jersey: John Wiley & Sons, Inc, 2005.
48. Temel, GO, Ankaralı, H. ve Yazıcı, AC (2010). Regresyon Modellerine Alternatif Bir Yaklaşım: MARS. *Türkiye Klinikleri Biyoistatistik Dergisi*, 2, 58-66.
49. Yücel, YB. Yaşam Memnuniyetini Etkileyen Faktörlerin Sınıflama ve Regresyon Ağacı ile Belirlenmesi. Yayınlanmamış Yüksek Lisans Tezi. İstanbul, İstanbul Ticaret Üniversitesi, FBE. 2017.
50. Irimia-Diequez, A., Blanco-Oliver, A., Vazquez-Cueto, M. (2015). “A Comparison of Classification/Regression Trees and Logistic Regression in Failure Models”. *Procedia Economics and Finance* 26, 23-28.
51. Breiman L. ve, Classification and Regression Trees, New York: Chapman & Hall, 1993, s.21.
52. Timofeew R., “Classification and Regression Trees (CART) Theory and Applications”, 2004, Humbolt University, Center of Applied Statistics and Economics,
53. Antipov E., Pokryshevskaya A, “Applying CHAID for logistic regression diagnostics and classification accuracy improvement”, *Journal of Targeting, Measurement and Analysis for Marketing*, 2010, 18:2, s.111

54. James G., Witten D, Hastie T, Tibshirani R, An Introduction to Statistical Learning, Springer. 2013, 316-321.
55. Tin Kam N. Random Decision Forests. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, 1995, 278-282.
56. Ho TK. "The Random Subspace Method for Constructing Decision Forests" Transactions on Pattern Analysis and Machine Intelligence. 1998, 20: 832-844.
57. Hastie, T. Tibshirani, Friedman, J. The Elements of Statistical Learning (2nd ed.). Springer, 2008, 587–588.
58. Pirayonesi S. Madeh; El-Diraby Tamer E. (2020). "Role of Data Analytics in Infrastructure Asset Management: Overcoming Data Size and Quality Problems". Journal of Transportation Engineering, Part B: Pavements. 146.
59. Pirayonesi, S. Madeh; El-Diraby, Tamer E. "Using Machine Learning to Examine Impact of Type of Performance Indicator on Flexible Pavement Deterioration Modeling" *Journal of Infrastructure Systems*. 2021, 27.
60. Kleinberg E "Stochastic Discrimination" Annals of Mathematics and Artificial Intelligence. 1990, 1: 207-239.
61. Kleinberg E. "An Overtraining-Resistant Stochastic Modeling Method for Pattern Recognition" Annals of Statistics, 1996. 24: 2319-2349.
62. Kleinberg E "On the Algorithmic Implementation of Stochastic Discrimination" IEEE Transactions on PAMI. 2000. 22: 47-90.
63. Breiman L. "Random forests" Machine Learning. 2001. 45: 5-32.
64. Sagi, Omer; Rokach, Lior. "Explainable decision forest: Transforming a decision forest into an interpretable tree. Information Fusion. 2020, 61: 124-138.
65. Liaw A. "Documentation for R package randomForest" <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf> 2012.

66. Trademark of Health Care Productivity, Inc. - Registration Number 3185828 - Serial Number 78642027, Justia Trademarks"
67. Archer, K. J. Emprical characterization of rastgele orman variable importance measure, computational statistical data analysis, Computational Statistics & Data Analysis, 2008, 52, 2249-2260.
68. Akar Ö., Harita Mühendisliği Anabilim Dalı Rastgele Orman Sınıflandırıcısına Doku Özellikleri Entegre Edilerek Benzer Spektral Özellikteki Tarımsal Ürünlerin Sınıflandırılması, Doktora Tezi, Trabzon, Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü, 2013.
69. Prasad, A.M., Iverson L.R. , Liaw A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction, Ecosystems, 2006, 9:181-199.
70. Akman, M. Genç, Y. Ankaralı, H. Rastgele ormans yöntemi ve sağlık alanında bir uygulama. Türkiye Klinikleri Journal of Biostatistics, 2011, 3: 36-48.
71. Tuğcu A, Aytekin S. Stres Ekokardiyografi. In Erol Ç, ed. Klinik Ekokardiyografi ve Diğer Görüntüleme yöntemleri. Ankara: Nobel Tıp Kitabevi, 2007;113-60
72. Feigenbaum H, Armstrong WF, Ryan T. *Feigenbaum's Echocardiography* Sixth Edition. Philadelphia: Lippincott Williams &Wilkins 2007;488-522.
73. Sicari R, Nihoyannopoulos P, Evangelista A. Stress echocardiography expert consensus statement (EAE). *Eur J of Echocardiography*, 2008;9:415-37
74. Tuttle RR, Mills J. Dobutamine: Development of new cathecolamine to selectively increase cardiac contractility. *Cirs Res*, 1975;36:185-96
75. Krahwinkel W, Ketteler T, Godke J, et al. Dobutamine stress echocardiography. *European Heart Journal* 1997;18:9-15

76. Leiner CV, Unverferth DV. Drugs five years later: dobutamine. *Ann Intern Med* 1983; 99:490-96
77. Chatterjee K, De Marco T. Central and peripheral adrenergic receptor agonists in heart failure. *European Heart Journal*, 1989;10:55-60
78. Fung AY, Gallagher GP, Buda AJ. Physiologic basis of dobutamine as compared with dipyridamole stress interventions in the assessment of critical coronary stenosis. *Circulation* 1987;76:943-51
79. Meyer SL, Curry GC, Donsky MS, et al. Influence of dobutamine on hemodynamics and coronary blood flow in patients with and without coronary artery disease. *Am J Cardiol*, 1976;38:103-8
80. Senior R, Lahiri A. Enhanced detection of myocardial ischemia by stress dobutamine echocardiography utilizing the "biphasic" response of wall thickening during low and high dose dobutamine infusion. *J Am Coll Cardiol*, 1995;26:26-31
81. Cerqueira MD. Imaging Techniques in Nuclear Cardiology. In Topol EJ ed. *Textbook of Cardiovascular Medicine* [3rd ed.]. Philadelphia. Lipincott Williams & Wilkins 2007;897-930
82. Acar E. St Elevasyonlu Miyokard İnfarktüsü Geçiren Hastalarda Dobutamin Stres Ekokardiyografi İle Canlılık Araştırmasında Speckle Tracking Ekokardiyografinin Yeri, Doktora Tezi, Kocaeli Üniversitesi Tıp Fakültesi, 2011.
83. Armstrong WF, Zoghbi WA. Stress echocardiography: Current methodology and clinical applications. *J Am Coll Cardiol*, 2005;45:1739-47
84. Orsinelli DA, Daniels CJ. Pharmacologic stress echocardiography. *Cardiology Clinics Of North America*, 1999;17:461-79
85. Gottdiener JS. Overview of stress echocardiography :Uses advantages, and limitations. *Curr Probl Cardiol*, 2003;8:485-512

86. Picano E. Stress echocardiography: From pathophysiological toy to diagnostic tool. *Circulation* 1992;85:1604-09
87. Krivokapich, J., Child, JS., Walter, DO., Garfinkel,A., Prognostic Value of Dobutamine Stress Echocardiography in Predicting Cardiac Events in Patients With Known or Suspected Coronary Artery Disease, *Journal of the American College of Cardiology*, 33, No. 3, 1999,ss. 708-716
88. (<https://cran.r-project.org/web/packages/vip/vignettes/vip-introduction.pdf>)
89. Xaplanteris P, Fournier S, Pijls NHJ, Fearon WF, Barbato E, Tonino PAL, Engstrøm T, Käåb S, Dambrink JH, Rioufol G, Toth GG, Piroth Z, Witt N, Fröbert O, Kala P, Linke A, Jagic N, Mates M, Mavromatis K, Samady H, Irimpen A, Oldroyd K, Campo G, Rothenbühler M, Jüni P, De Bruyne B; FAME 2 Investigators. Five-Year Outcomes with PCI Guided by Fractional Flow Reserve. *N Engl J Med*. 2018 Jul 19;379(3):250-259.
90. Wanner C, Inzucchi SE, Lachin JM, Fitchett D, von Eynatten M, Mattheus M, Johansen OE, Woerle HJ, Broedl UC, Zinman B; EMPA-REG OUTCOME Investigators. Empagliflozin and Progression of Kidney Disease in Type 2 Diabetes. *N Engl J Med*. 2016 Jul 28;375:323-34.. PMID:
91. Ridderstråle M, Andersen KR, Zeller C, Kim G, Woerle HJ, Broedl UC; EMPA-REG H2H-SU trial investigators. Comparison of empagliflozin and glimepiride as add-on to metformin in patients with type 2 diabetes: a 104-week randomised, active-controlled, double-blind, phase 3 trial. *Lancet Diabetes Endocrinol*. 2014 Sep;2(9):691-700.
92. Van Mieghem NM, Deeb GM, Søndergaard L, Grube E, Windecker S, Gada H, Mumtaz M, Olsen PS, Heiser JC, Merhi W, Kleiman NS, Chetcuti SJ, Gleason TG, Lee JS, Cheng W, Makkar RR, Crestanello J, George B, George I, Kodali S,

- Yakubov SJ, Serruys PW, Lange R, Piazza N, Williams MR, Oh JK, Adams DH, Li S, Reardon MJ; SURTAVI Trial Investigators. Self-expanding Transcatheter vs Surgical Aortic Valve Replacement in Intermediate-Risk Patients: 5-Year Outcomes of the SURTAVI Randomized Clinical Trial. *JAMA Cardiol.* 2022 Oct 1;7(10):1000-1008.
93. Dolan G, Smith LA, Collins S, Plumb JM. Effect of setting, monitoring intensity and patient experience on anticoagulation control: a systematic review and meta-analysis of the literature. *Curr Med Res Opin.* 2008 May;24(5):1459-72.
94. Canto JG, Every NR, Magid DJ, Rogers WJ, Malmgren JA, Frederick PD, French WJ, Tiefenbrunn AJ, Misra VK, Kiefe CI, Barron HV. The volume of primary angioplasty procedures and survival after acute myocardial infarction. National Registry of Myocardial Infarction 2 Investigators. *N Engl J Med.* 2000 May 25;342(21):1573-80.
95. Montalescot G, White HD, Gallo R, Cohen M, Steg PG, Aylward PE, Bode C, Chiariello M, King SB 3rd, Harrington RA, Desmet WJ, Macaya C, Steinhubl SR; STEEPLE Investigators. Enoxaparin versus unfractionated heparin in elective percutaneous coronary intervention. *N Engl J Med.* 2006 Sep 7;355(10):1006-17.
96. Peacock WF 4th, De Marco T, Fonarow GC, Diercks D, Wynne J, Apple FS, Wu AH; ADHERE Investigators. Cardiac troponin and outcome in acute heart failure. *N Engl J Med.* 2008 May 15;358(20):2117-26.
97. Pilote L, Miller DP, Califf RM, Rao JS, Weaver WD, Topol EJ. Determinants of the use of coronary angiography and revascularization after thrombolysis for acute myocardial infarction. *N Engl J Med.* 1996 Oct 17;335(16):1198-205.
98. REMAP-CAP Investigators; ACTIV-4a Investigators; ATTACC Investigators; Goligher EC, Bradbury CA, McVerry BJ at all. Therapeutic Anticoagulation with

- Heparin in Critically Ill Patients with Covid-19. *N Engl J Med*. 2021 Aug 26;385(9):777-789.
99. Mancia G, Rea F, Ludergnani M, Apolone G, Corrao G. Renin-Angiotensin-Aldosterone System Blockers and the Risk of Covid-19. *N Engl J Med*. 2020 Jun 18;382(25):2431-2440.
100. Venepally NR, Arsanjani R, Agasthi P, Wang P, Khetarpal BK, Barry T, Chao CJ, Fath AR, Mookadam F. A New Insight Into Nonbacterial Thrombotic Endocarditis: A Systematic Review of Cases. *Anatol J Cardiol*. 2022 Oct;26(10):743-749.
101. Walther CP, Benoit JS, Gregg LP, Bansal N, Nambi V, Feldman HI, Shlipak MG, Navaneethan SD; CRIC Study Investigators. Heart failure-type symptom scores in chronic kidney disease: The importance of body mass index. *Int J Obes (Lond)*. 2022 Oct;46(10):1910-1917.

EKLER

EK-1. ÖZGEÇMİŞ

Kişisel Bilgiler
Adı Soyadı: Doğum tarihi: Doğum Yeri: Medeni Hali: Uyruğu: Adres: Tel: Faks: E-mail:
Eğitim
Lise: Lisans: Tıpta uzmanlık:
Yabancı Dil Bilgisi
İngilizce: Almanca: - Rusça: -
Üye Olunan Mesleki Kuruluşlar
İlgi Alanları ve Hobiler

EK-2. ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU



SAĞLIK BİLİMLERİ ENSTİTÜSÜ
Graduate School of Health Sciences

ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU

Öğrencinin Adı ve Soyadı	SELİM TOPCU
Öğrencinin Numarası	
Ana Bilim Dalı	Biyostatistik ve Tıp Bilimleri
Öğrencinin Kayıtlı Olduğu Program Türü	Doktora

Yukarıda bilgileri verilen tezin intihal tespit yazılımıyla (Turnitin) yapılan tarama sonucunda elde edilen benzerlik oranları aşağıdaki gibidir. Beyan edilen bilgilerin doğru olduğunu, aksi hâlde doğacak hukuki sorumlulukları kabul ve beyan ederiz.

Bölümler	Benzerlik Oranı	Maksimum Benzerlik Oranları
I. Giriş	%1	% 15
II. Genel Bilgiler	%29	% 35
III. Materyal ve Metod	%14	% 35
IV. Bulgular	%1	% 15
V. Tartışma	%1	% 20

Not: Yedi kelimeye kadar benzerlikler ile Başlık, Kaynakça, İçindekiler, Teşekkür, Dizin ve Ekler kısımları tarama dışı bırakılabilir. Yukarıdaki azami benzerlik oranları yanında tek bir kaynaktan olan benzerlik oranlarının %5’den büyük olmaması gerekir.

Tez Yazarı (Öğrenci)

Tez Danışmanı

EK-3. ETİK KURUL ONAY FORMU



**ATATÜRK ÜNİVERSİTESİ TIP
FAKÜLTESİ KLİNİK ARAŞTIRMALAR
ETİK KURULU**



ETİK KURUL BİLGİLERİ		KARAR	
ETİK KURULUN ADI	Atatürk Üniversitesi Tıp Fakültesi Klinik Araştırmalar Etik Kurulu		
AÇIK ADRESİ:	Atatürk Üniversitesi Tıp Fakültesi Dekanlığı		
TELEFON	+90 442 234 65 11		
FAKS	+90 442 236 09 68		
E-POSTA	atatipetikkurul@gmail.com		
SORUMLU ARAŞTIRMACI UNVANI/ADI/SOYADI	Doç.Dr.Selim TOPÇU		
ARAŞTIRMACININ AÇIK ADI	Çok Değişkenli Regresyon Ağaçları ve Kardiyoloji Verilerine Uygulanması		
KARAR BİLGİLERİ	Toplantı Sayısı: 04 Karar No: 100	Tarih: 30.05.2019	
	Yukarıda bilgileri verilen başvuru dosyası ile ilgili belgeler araştırmanın/çalışmanın gerekçe, amaç, yaklaşım ve yöntemleri dikkate alınarak incelenmiş ve çalışmanın bütçesinin kendisi tarafından karşılanması koşulu ile yapılmasında bilimsel ve etik açıdan sakınca olmadığına oy birliği ile karar verildi. Klinik Araştırmalar Hakkında Yönetmelik kapsamında yer alan araştırmalar/çalışmalar için Türkiye İlaç ve Tıbbi Cihaz Kurumu'ndan izin alınması gerekmektedir. Araştırmacıya çalışmalarında başarılar dileriz.		

TEMEL TIP BİLİMLERİ BÖLÜM KURUL KARARI

TARİH: 10.6.2019

NO: 7

Temel Tıp Bilimleri Bölüm Kurulu Prof.Dr. Ebubekir BAKAN'ın başkanlığında toplanarak aşağıdaki kararları almıştır.

Karar 1-Biyostatistik ve Tıp Bilimi Anabilim Dalı Araştırma Görevlilerinden Dr. Selim TOPCU'ya "Çok Değişkenli Regresyon Ağaçları ve Kardiyoloji Verilerine Uygulanması" isimli çalışmanın tez konusu olarak verilmesine,

Karar No 2- Danışmanlığını Prof.Dr. Suphi ÖZÇOMAK'ın yürütmesine,

oy birliği ile karar verilmiştir.



EK-4. TEZ ADI DEĞİŞİKLİĞİ FORMU



SAĞLIK BİLİMLERİ ENSTİTÜSÜ
Graduate School of Health Sciences

TEZ ADI/KONUSU DEĞİŞİKLİĞİ BİLDİRİM FORMU

Öğrencinin Adı ve Soyadı	Selim TOPCU
Ana Bilim Dalı	Biyoistatistik ve Tıp Bilişimi
Öğrencinin Kayıtlı Olduğu Program Türü	Doktora

Biyoistatistik ve Tıp Bilişimi Ana Bilim Dalı Başkanlığına

Danışmanlığımı yürüttüğüm ve yukarıda bilgileri yazılı olan bilim dalımız öğrencisinin Tez adı aşağıda belirtilen şekilde değiştirilmiştir. Bilgilerinize arz ederim. 10.02.2023

Değişiklik Türü	Tez Adı Değişikliği ☒ Tez İçeriği Değişikliği ¹ ☐
Tezin Eski Adı	Çok Değişkenli Regresyon Ağaçları Ve Kardiyoloji Verilerine Uygulanması
Tezin Yeni Adı	Koroner Arter Hastalığında Klinik Sonlanımların Öngörülmesinde Lojistik Regresyon Analizi, CART ve Rastgele Orman Algoritmalarının Karşılaştırılması
Tezin İngilizce Adı	Comparison of Logistic Regression Analysis, CART and Random Forest Algorithms for Predicting Clinical Outcomes in Patients with Coronary Artery Disease
Değişikliğin Gerekçesi	Tezin uygulama bölümünde koroner arter hastalarından elde edilen bulguların kullanılması.

Tez İzleme Komitesi Üyeleri	İmza
-----------------------------	------